

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 22:30:21 ; Search time 973.268 Seconds
(without alignments)
8940.745 Million cell updates/sec

Title: US-10-029-359A-1

Perfect score: 299

Sequence: 1 gcaagaggagcagccgcg.....tcgctgaaacacctctattag 299

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:**

- 1: gb_ba:**
- 2: gb_htg:**
- 3: gb_in:**
- 4: gb_cm:**
- 5: gb_ov:**
- 6: gb_pat:**
- 7: gb_ph:**
- 8: gb_pl:**
- 9: gb_pr:**
- 10: gb_ro:**
- 11: gb_sts:**
- 12: gb_sy:**
- 13: gb_un:**
- 14: gb_vi:**
- 15: em_ba:**
- 16: em_fun:**
- 17: em_hum:**
- 18: em_in:**
- 19: em_mu:**
- 20: em_om:**
- 21: em_or:**
- 22: em_ov:**
- 23: em_pat:**
- 24: em_ph:**
- 25: em_pl:**
- 26: em_ro:**
- 27: em_sts:**
- 28: em_un:**
- 29: em_vi:**
- 30: em_htg_hum:**
- 31: em_htg_inv:**
- 32: em_htg_other:**
- 33: em_htg_mus:**
- 34: em_htg_pln:**
- 35: em_htg_rod:**
- 36: em_htg_mam:**
- 37: em_htg_vrt:**
- 38: em_sy:**
- 39: em_hgtc_hum:**
- 40: em_hgtc_mus:**
- 41: em_hgtc_other:**

Pred. NO. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	283.8	94.9	156337	2	AC009938 Homo sapi
C 2	283.8	94.9	169898	2	AC016585 Homo sapi
C 3	283.8	94.9	172574	2	AC011121 Homo sapi
C 4	283.8	94.9	162823	9	AC016638 Homo sapi
C 5	283.8	94.9	217807	2	AC079469 Homo sapi
C 6	283.8	94.9	1794	6	AX175160 Sequence
C 7	236.4	79.1	1794	9	AF155582 Homo sapi
C 8	231	77.3	930	6	HSJ132443 Homo sapi
C 9	231	77.3	1092	6	AX175170 Sequence
C 10	231	77.3	1092	9	HA243256 Homo sapi
C 11	231	77.3	1092	9	HA278960 Homo sapi
C 12	206.6	69.1	115954	9	AC005532 Homo sapi
C 13	194	64.9	1438	10	AF157962 Mus muscu
C 14	194	64.9	1469	6	AX175164 Sequence
C 15	192.6	64.4	1092	6	AX175172 Sequence
C 16	190.4	63.7	123070	9	AC020656 Homo sapi
C 17	190.4	63.7	307755	2	AC127894 Homo sapi
C 18	189.2	63.3	1405	10	AF157963 Rattus no
C 19	189.2	63.3	1440	6	AX175162 Sequence
C 20	188.8	63.1	170227	2	AC019161 Homo sapi
C 21	187.8	62.8	1092	6	AX175171 Sequence
C 22	169.8	56.8	185872	2	AC068810 Mus muscu
C 23	163.4	54.6	194038	2	AC095494 Rattus no
C 24	145.8	48.8	171529	2	AC023021 Homo sapi
C 25	111.4	37.3	146790	2	AC118920 Rattus no
C 26	83.4	27.9	167996	9	AC021753 Homo sapi
C 27	83.4	27.9	191655	9	AC020661 Homo sapi
C 28	78.8	26.4	153472	9	AC011179 Homo sapi
C 29	78.8	26.4	165720	9	AC025524 Homo sapi
C 30	67.2	22.5	183416	10	AL671269 Mouse DNA
C 31	67.2	22.5	229117	2	AL807383 Mus muscu
C 32	53.2	17.8	1708	10	BC025899 Mus muscu
C 33	50.8	17.0	121200	2	AC112336 Rattus no
C 34	42	14.0	213308	2	AC079164 Mus muscu
C 35	42	14.0	217953	2	AC122334 Mus muscu
C 36	41.6	13.9	173665	2	AC115770 Mus muscu
C 37	40.8	13.6	187431	9	AC107219 Homo sapi
C 38	40.4	13.5	68573	2	AC107791 Mus muscu
C 39	40.2	13.4	176113	2	AC103226 Rattus no
C 40	40.2	13.4	191821	2	AC123475 Rattus no
C 41	40	13.4	173430	2	AC112689 Mus muscu
C 42	39.8	13.3	167274	2	AC114246 Rattus no
C 43	39.2	13.1	3322	8	R1CFUJI D50602 Rice mRNA f
C 44	39	13.0	164581	2	AC126794 Medicago
C 45	38.8	13.0	148950	2	AC108583 Rattus no

ALIGNMENTS

RESULT 1
AC009938/c 156337 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-115B14, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC009938
ACCESSION AC009938
VERSION AC009938.3 GI:9113959
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 156337)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-115B14
JOURNAL Unpublished

REFERENCE
AUTHORS

2 (bases 1 to 156337)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,
 Castle, A., Cerny, J., Collangelo, M., Collins, S., Collymore, A.,
 Cooke, P., Deaillano, K., Depayre, E., Devon, K., Dewar, K.,
 Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,
 Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
 Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
 Karas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,
 Marquis, N., McEwan, P., McCurk, A., McKernan, K., McLaughlin, J.,
 Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
 Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
 Testaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
 Direct Submission
 Submitted (08-SEP-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 13, 2000 this sequence version replaced gi:6978143.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L1978
 Center clone name: L15_B_14

* NOTE: This record contains 149 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 995: contig of 995 bp in length
 * 996 1095: gap of 100 bp
 * 1096 2065: contig of 970 bp in length
 * 2066 2165: gap of 100 bp
 * 2166 3117: contig of 952 bp in length
 * 3118 3217: gap of 100 bp
 * 3218 4168: contig of 951 bp in length
 * 4169 4268: gap of 100 bp
 * 4269 5236: contig of 968 bp in length
 * 5237 5336: gap of 100 bp
 * 5337 6309: contig of 973 bp in length
 * 6310 6409: gap of 100 bp
 * 6410 7365: contig of 956 bp in length
 * 7366 7465: gap of 100 bp
 * 7466 8417: contig of 952 bp in length
 * 8418 8517: gap of 100 bp
 * 8518 9471: contig of 954 bp in length
 * 9472 9571: gap of 100 bp
 * 9572 10511: contig of 940 bp in length
 * 10512 10611: gap of 100 bp
 * 10612 11556: contig of 945 bp in length
 * 11557 11656: gap of 100 bp
 * 11657 12627: contig of 971 bp in length
 * 12628 12727: gap of 100 bp
 * 12728 13655: contig of 928 bp in length
 * 13656 13755: gap of 100 bp
 * 13756 14734: contig of 979 bp in length
 * 14735 14834: gap of 100 bp
 * 14835 15782: contig of 948 bp in length
 * 15783 15882: gap of 100 bp

15883 16812: contig of 930 bp in length
 * 16813 16912: gap of 100 bp
 * 16913 17855: contig of 943 bp in length
 * 17856 17955: gap of 100 bp
 * 17956 18939: contig of 984 bp in length
 * 18940 19039: gap of 100 bp
 * 19040 19995: contig of 956 bp in length
 * 19996 20095: gap of 100 bp
 * 20096 21022: contig of 927 bp in length
 * 21023 21122: gap of 100 bp
 * 21123 22053: contig of 931 bp in length
 * 22054 22153: gap of 100 bp
 * 22154 23146: contig of 993 bp in length
 * 23147 23246: gap of 100 bp
 * 23247 24206: contig of 960 bp in length
 * 24207 24306: gap of 100 bp
 * 24307 25275: contig of 969 bp in length
 * 25276 25375: gap of 100 bp
 * 25376 26338: contig of 963 bp in length
 * 26339 26438: gap of 100 bp
 * 26439 27420: contig of 982 bp in length
 * 27421 27520: gap of 100 bp
 * 27521 28483: contig of 963 bp in length
 * 28484 28584: gap of 100 bp
 * 28584 29532: contig of 949 bp in length
 * 29533 29632: gap of 100 bp
 * 29633 30571: contig of 939 bp in length
 * 30572 30671: gap of 100 bp
 * 30672 31631: contig of 960 bp in length
 * 31632 31731: gap of 100 bp
 * 31732 32674: contig of 943 bp in length
 * 32675 32774: gap of 100 bp
 * 32775 33735: contig of 961 bp in length
 * 33736 33835: gap of 100 bp
 * 33836 34756: contig of 921 bp in length
 * 34757 34856: gap of 100 bp
 * 34857 35784: contig of 928 bp in length
 * 35785 35884: gap of 100 bp
 * 35885 36818: contig of 934 bp in length
 * 36819 36918: gap of 100 bp
 * 36919 37873: contig of 955 bp in length
 * 37874 37973: gap of 100 bp
 * 37974 38939: contig of 966 bp in length
 * 38940 39039: gap of 100 bp
 * 39040 39953: contig of 914 bp in length
 * 39954 40053: gap of 100 bp
 * 40054 40999: contig of 946 bp in length
 * 41000 41099: gap of 100 bp
 * 41100 42004: contig of 905 bp in length
 * 42005 42104: gap of 100 bp
 * 42105 43018: contig of 914 bp in length
 * 43019 43118: gap of 100 bp
 * 43119 44057: contig of 939 bp in length
 * 44058 44157: gap of 100 bp
 * 44158 45127: contig of 970 bp in length
 * 45128 45227: gap of 100 bp
 * 45228 46166: contig of 939 bp in length
 * 46167 46266: gap of 100 bp
 * 46267 47241: contig of 975 bp in length
 * 47242 47341: gap of 100 bp
 * 47342 48308: contig of 967 bp in length
 * 48309 48408: gap of 100 bp
 * 48409 49391: contig of 983 bp in length
 * 49392 49491: gap of 100 bp
 * 49492 50414: contig of 923 bp in length
 * 50415 50514: gap of 100 bp
 * 50515 51482: contig of 968 bp in length
 * 51483 51582: gap of 100 bp
 * 51583 52526: contig of 944 bp in length
 * 52527 52626: gap of 100 bp
 * 52627 53610: contig of 984 bp in length
 * 53611 53710: gap of 100 bp
 * 53711 54676: contig of 966 bp in length

* 54677 54776: gap of 100 bp
* 54777 55729: contig of 953 bp in length
* 55730 55829: gap of 100 bp
* 55830 56781: contig of 952 bp in length
* 56782 56881: gap of 100 bp
* 56882 57863: contig of 982 bp in length
* 57864 57963: gap of 100 bp
* 57964 58892: contig of 929 bp in length
* 58893 58992: gap of 100 bp
* 58993 59960: contig of 968 bp in length
* 59961 60060: gap of 100 bp
* 60061 61030: contig of 970 bp in length
* 61031 61130: gap of 100 bp
* 61131 62075: contig of 945 bp in length
* 62076 62175: gap of 100 bp
* 62176 63111: contig of 936 bp in length
* 63112 63211: gap of 100 bp
* 63212 64148: contig of 937 bp in length
* 64149 64248: gap of 100 bp
* 64249 65186: contig of 938 bp in length
* 65187 65286: gap of 100 bp
* 65287 66257: contig of 971 bp in length
* 66258 66357: gap of 100 bp
* 66358 67306: contig of 949 bp in length
* 67307 67406: gap of 100 bp
* 67407 68325: contig of 919 bp in length
* 68326 68425: gap of 100 bp
* 68426 69370: contig of 945 bp in length
* 69371 69470: gap of 100 bp
* 69471 70457: contig of 987 bp in length
* 70458 70557: gap of 100 bp
* 70558 71497: contig of 940 bp in length
* 71498 71597: gap of 100 bp
* 71598 72513: contig of 916 bp in length
* 72514 72613: gap of 100 bp
* 72614 73585: contig of 972 bp in length
* 73586 73685: gap of 100 bp
* 73686 74654: contig of 969 bp in length
* 74655 74754: gap of 100 bp
* 74755 75734: contig of 980 bp in length
* 75735 75834: gap of 100 bp

Query Match 94.9%; Score 283.8; DB 2; Length 156337;
Best Local Similarity 99.0%; Pred. No. 1.8e-63;
Matches 296; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GCAGAGGGAGCCAGCGCGGATGACAGAAATTCACCTTCCGAGATGGCTCTAAATCCT 60
Db 139022 GCAGAGGGAGCCAGCGCGGATGACAG-AAATACACTTCCGAGATGGCTCTAAATCCT 138964
QY 61 GCCTGAATTTTAACTTCCCTATGGATCGGCAATAGGGTTTATTTTATTTCTCAGC 120
Db 138963 GCCTGAATTTTAACTTCCCTATGGATCGGCAATAGGGTTTATTTTATTTCTCAGC 138904
QY 121 TACTTAGTATTTTGTGGGAGAAGAGGGTGCACCCAGACTAATGCTTCTCATATGATC 180
Db 138903 TACTTAGTATTTTGTGGGAGAAGAGGGTGCACCCAGACTAATGCTTCTCATATGATC 180
QY 181 CTCATCGAGGATTCAGATGATATGACAGATCATCTAGGAGGACAAATGAATCTCA 240
Db 138843 CTCATCGAGGATTCAGATGATATGACAGATCATCTAGGAGGACAAATGAATCTCA 138784
QY 241 ATGCAGATTCAGCCAACTGAAGATGAGAACAGAGAAATCGCTGAAACCTCTATTAG 299
Db 138783 ATGCAGATTCAGCCAACTGAAGATGAGAACAGAGAAATCGCTGAAACCTCTATTAG 138725

RESULT 2
AC016585/c
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTD-2601K17, WORKING DRAFT
SEQUENCE, 3 unordered pieces.
AC016585
AC016585

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC016585.6 GI:15290346.
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 169898)
DOE Joint Genome Institute.
Direct Submission
Submitted (04-DEC-1999), Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Aug 25, 2001 this sequence version replaced gi:13699586.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 829022
Center clone name: CITB-EL_2601K17

Summary Statistics
Consensus quality: 168838 bases at least Q40
Consensus quality: 169474 bases at least Q30
Consensus quality: 169606 bases at least Q20
Estimated insert size: 167000; pulse field gel estimation
Quality coverage: 7.44 in Q20 bases; sum-of-contigs estimation
Quality coverage: 7.32 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 10691: contig of 10691 bp in length
* 10692 10791: gap of unknown length
* 10792 53413: contig of 42622 bp in length
* 53414 53513: gap of unknown length
* 53514 169898: contig of 116385 bp in length.
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2601K17"
/clone_lib="CalTech human BAC library D"
BASE COUNT 49510 a 35344 c 35770 g 49074 t 200 others
ORIGIN

Query Match 94.9%; Score 283.8; DB 2; Length 169898;
Best Local Similarity 99.0%; Pred. No. 1.8e-63;
Matches 296; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GCAAGAGGGAGCCAGCGCGGATGACAGAAATTCACCTTCCGAGATGGCTCTAAATCCT 60
Db 112793 GCAAGAGGGAGCCAGCGCGGATGACAG-AAATACACTTCCGAGATGGCTCTAAATCCT 112735
QY 61 GCCTGAATTTTAACTTCCCTATGGATCGGCAATAGGGTTTATTTTATTTCTCAGC 120
Db 112734 GCCTGAATTTTAACTTCCCTATGGATCGGCAATAGGGTTTATTTTATTTCTCAGC 112675
QY 121 TACTTAGTATTTTGTGGGAGAAGAGGGTGCACCCAGACTAATGCTTCTCATATGATC 180
Db 112674 TACTTAGTATTTTGTGGGAGAAGAGGGTGCACCCAGACTAATGCTTCTCATATGATC 112615
QY 181 CTCATCGAGGATTCAGATGATATGACAGAAATTCAGAGGACAAATGAATCTCA 240

```

|||||
Db 112614 CTCATCGAGGCAATCAGATGATAATGGACAGAAATCATCTAGGAGGACAAATGAACCTCA 112555
|||||
QY 241 ATCCAGATCTACCCACGTAAGATGAGACACAGAAATCGTGAACCTCTATTAG 299
|||||
Db 112554 ATCCAGATCTACCCACGTAAGATGAGACACAGACATCGTGAACCTCTATTAG 112496
|||||

```

RESULT 3

```

AC011121/c
LOCUS AC011121.5 GI:9666267 linear HTG 04-SEP-2000
DEFINITION Homo sapiens clone RP11-364C6, WORKING DRAFT SEQUENCE, 10 unordered
pieces.
AC011121
VERSION AC011121.5
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 172574)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome, clone RP11-364C6
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 172574)

```

```

REFERENCE 1 (bases 1 to 172574)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
TITLE Homo sapiens chromosome, clone RP11-364C6
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 172574)
Balwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArlano, K., Dewar, K., Domino, M., Donegan, L., Doyle, M.,
Ferrel, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galligan, J., Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Hawland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 4, 2000 this sequence version replaced gi:7630654.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2174
Center clone name: 364_C6
----- Summary Statistics
Sequencing vector: M13; M77815; 96% of reads
Sequencing vector: Plasmid; n/a; 0.0% of reads
3.84985563041386Chemistry: Dye-terminator Big Dye; 100% of
reads
Assembly program: Phrap; version 0.960731
Consensus quality: 160604 bases at least Q40
Consensus quality: 165990 bases at least Q30
Consensus quality: 168767 bases at least Q20
Insert size: 182000; agarose-ff
Insert size: 171674; sum-of-ctgigs
Quality coverage: 4.5 in Q20 bases; agarose-ff
Quality cov.

```

* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence

```

* as soon as it is available and the accession number will
* be preserved
* 1 46314: contig of 46314 bp in length
* 46315 46414: gap of 100 bp
* 46415 48772: contig of 2358 bp in length
* 48773 48872: gap of 100 bp
* 48873 51876: contig of 3004 bp in length
* 51877 51976: gap of 100 bp
* 51977 54631: contig of 2655 bp in length
* 54632 54731: gap of 100 bp
* 54732 59065: contig of 4334 bp in length
* 59066 59165: gap of 100 bp
* 59166 63397: contig of 4232 bp in length
* 63398 63497: gap of 100 bp
* 63498 117527: contig of 54030 bp in length
* 117528 117627: gap of 100 bp
* 117628 143391: contig of 25764 bp in length
* 143392 143491: gap of 100 bp
* 143492 167784: contig of 24293 bp in length
* 167785 167884: gap of 100 bp
* 167885 172574: contig of 4690 bp in length.

```

FEATURES

source

```

1..172574
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP11-364C6"

```

```

misc_feature
1..46314
/clone="assembly_fragment"
/clone_end:SP6
vector_side:left"
46415..48772
/notes="assembly_fragment"
48873..51876
/notes="assembly_fragment"
51977..54631
/notes="assembly_fragment"
54732..59065
/notes="assembly_fragment"
59166..63397
/notes="assembly_fragment"
63498..117527
/notes="assembly_fragment"
117628..143391
/notes="assembly_fragment"
143492..167784
/notes="assembly_fragment"
167885..172574
/notes="assembly_fragment"
clone_end:T7
vector_side:right"

```

```

BASE COUNT 50796 a 35214 c 36052 g 49603 t 909 others
ORIGIN

```

```

Query Match 94.9%; Score 283.8; DB 2; Length 172574;
Best Local Similarity 99.0%; Pred. No. 1.8e-63;
Matches 296; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```

```

QY 1 GCAAGAGGAGGACGCGCGATGACAGAAAATTCACATCCGAGATGGCCTCTAAATCCT 60
DB 80478 GCAAGAGGAGGACGCGCGATGACAG-AAAATACACTTTCCGAGATGGCCTCTAAATCCT 80420
QY 61 GCGTGAATTTTAACTTCTCTATGGATCGGCAATAGGGTTATTTTATTTCTCAGC 120
DB 80419 GCGTGAATTTTAACTTCTCTATGGATCGGCAATAGGGTTATTTTATTTCTCAGC 80360
QY 121 TACTTAGTATTTTGGTGGAGAGGAGGTGACACCCAGACTAATGTCTTTCATATGATC 180
DB 80359 TACTTAGTATTTTGGTGGAGAGGAGGTGACACCCAGACTAATGTCTTTCATATGATC 80300
QY 181 CTCATCGGAGGCAATTCAGATGATATGGACAGAAATCATCTAGGAGGACAAATGAATTC 240
DB 80299 CTCATCGGAGGCAATTCAGATGATATGGAGAGAAATCATCTAGGAGGACAAATGAATTC 80240

```

```
QY 241 ATGCAGATTCTAGCCACGCTAAGATGAGAACACAGAAATCGCTGAAAACCTCTATTAG 299
|||||
Db 80239 ATGCAGATTCTAGCCACGCTAAGATGAGAACACAGACATCGCTGAAAACCTCTATTAG 80181
|||||

RESULT 4
AC016638/c 182823 bp DNA linear PRI 21-JUL-2001
LOCUS Homo sapiens chromosome 5 clone RP11-365H8, complete sequence.
DEFINITION AC016638
ACCESSION AC016638.9 GI:14993683
VERSION
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 21, 2001 this sequence version replaced gi:14861720.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.5% of Sequence;
Estimated Total Number of Errors is 0.7.
STS Content:
SHGC-3092 G16892
SHGC-18259 G31915
WI-4513 G02898.
Location/Qualifiers
1. 182823
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-365H8"
BASE COUNT 53032 a 39003 c 37540 g 53248 t
ORIGIN
Query Match 94.9%; Score 283.8; DB 9; Length 182823;
Best Local Similarity 99.0%; Pred. No. 1.8e-63;
Matches 296; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 GCAGAGGGAGCCAGCGCGATGACAGAAAATTCACTTCCGAGATGCCCTCTAAATCCT 60
|||||
Db 121704 GCAGAGGGAGCCAGCGCGATGACAG-AAAATACATTTCCGAGATGCCCTCTAAATCCT 121646
|||||
QY 61 GCCTCAATTTTAACTTCTCTATGATCGGCAATAGGGTTATTTATTTCTCTCAGC 120
|||||
Db 121645 GCCTCAATTTTAACTTCTCTATGATCGGCAATAGGGTTATTTATTTCTCTCAGC 121586
|||||
QY 121 TACTTAGTATTTTGTGGGAGAGAGGTGACACCCAGACTAAATGTTCTTCATATGATC 180
|||||
Db 121585 TACTTAGTATTTTGTGGGAGAGAGGTGACACCCAGACTAAATGTTCTTCATATGATC 121526
|||||
QY 181 CTCATCGGAGGCATTCAGATGATATGAGACAGATCATCTAGGAGGACAAATGAACITCA 240
|||||
```

```
Db 121525 CTCATCGGAGGCATTCAGATGATATGAGACAGATCATCTAGGAGGACAAATGAACITCA 121466
|||||
QY 241 ATGCAGATTCTAGCCACGCTAAGATGAGAACACAGAAATCGCTGAAAACCTCTATTAG 299
|||||
Db 121465 ATGCAGATTCTAGCCACGCTAAGATGAGAACACAGACATCGCTGAAAACCTCTATTAG 121407
|||||

RESULT 5
AC079469 217807 bp DNA linear HTG 19-APR-2001
LOCUS Homo sapiens chromosome 5 clone RP11-490H18, WORKING DRAFT
DEFINITION AC079469
ACCESSION AC079469.2 GI:13677063
VERSION
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 19, 2001 this sequence version replaced gi:9964834.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 594711
Center clone name: RPCI-11_490H18
-----
Summary Statistics
Consensus quality: 213920 bases at least Q40
Consensus quality: 216734 bases at least Q30
Consensus quality: 217279 bases at least Q20
Estimated insert size: 196000; pulse field gel estimation
Estimated insert size: 217507; sum-of-contigs estimation
Quality coverage: 8.1 in Q20 bases; pulse field gel estimation
Quality coverage: 7.37 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 28216: contig of 28216 bp in length
* 28217 28316: gap of unknown length
* 28317 41458: contig of 13142 bp in length
* 41459 41558: gap of unknown length
* 41559 155568: contig of 114010 bp in length
* 155569 155668: gap of unknown length
* 155669 217807: contig of 62139 bp in length.
Location/Qualifiers
1. 217807
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-490H18"
BASE COUNT 63364 a 45033 c 45462 g 63648 t 300 others
ORIGIN
```

```
Query Match          94.9%; Score 283.8; DB 2; Length 217807;
Best Local Similarity 99.0%; Pred. No. 1.8e-63;
Matches 296; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GCAGAGGGAGCCACGCCCGATGACAGAAATTCACATTTCCGAGATGGCCCTCTAAATCCT 60
DB 10597 GCAGAGGGAGCCACGCCCGATGACAG-AAATACATTTCCGAGATGGCCCTCTAAATCCT 10655
QY 61 GGCTGAATTTTAAACCTTCTCTATGATCGGCAATAGGTTTATTTATTTTCTCAGC 120
DB 10656 GGCTGAATTTTAAACCTTCTCTATGATCGGCAATAGGTTTATTTATTTTCTCAGC 10715
QY 121 TACTAGTATTTTGGGAGAGAGGGTGCACCCAGACTAATGTTCTTCATAATGATC 180
DB 10716 TACTAGTATTTTGGGAGAGAGGGTGCACCCAGACTAATGTTCTTCATAATGATC 10775
QY 181 CTATCGGAGGATTCAGATGATAATGGACAGATCATCTAGGAGGACAAATGAATCTCA 240
DB 10776 CTATCGGAGGATTCAGATGATAATGGACAGATCATCTAGGAGGACAAATGAATCTCA 10835
QY 241 ATCGAGATTCAGCCACGTAAGATGAGAACACAGAAATCGCTGAAACCTCTATTAG 299
DB 10836 ATCGAGATTCAGCCACGTAAGATGAGAACACAGACATCGCTGAAACCTCTATTAG 10894

RESULT 6
AX175160 1794 bp DNA linear PAT 03-JUL-2001
LOCUS Sequence 2 from Patent WO0144478.
ACCESSION AX175160
VERSION AX175160.1 GI:14598562
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1794)
AUTHORS Canfield,W.M., Cummings,R.D. and Ju,T.
TITLE Core lb3-galactosyl transferases and methods of use thereof
JOURNAL Patent: WO 0144478-A 21-JUN-2001;
The Board of Regents of The University of Oklahoma (US)
FEATURES
source 1..1794
/organism="Homo sapiens"
/db_xref="taxon:9606"
63..1154
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAC43870.1"
/db_xref="GI:14598563"
/translation="MASKSWLNFLTFLCGSAIGFLCGLSOLFLLGKVDTPNVLHN
DPHARSDGONHLEGOMNFADSSOHKQDENTDIAENLYOKVRIILCWMTGPNLEK
KAKHVATWAQRCNKVLFWSSEENKDFPAVLGKTEGRQLYWKTKAQYVHEHYLE
DADWFLKADDDTYVILDNRLWLLSKYDPEPIYFGRFRFPYVKQYMSGGAGYVLSKE
AKRFYDAFKTKDTHSSIEDIALGRCEIMNVEAGDSRDITGKTFHPFVPEHHLI
KGYLPRFWNYYPVVEGPGCCSDLAVSFHYVDSVTMYELEYLVYHLRYPGYLYR
YOPTLPERILKEISOANKNEDTKVKLGNP"
BASE COUNT 624 a 300 c 334 g 536 t
ORIGIN

Query Match          79.1%; Score 236.4; DB 6; Length 1794;
Best Local Similarity 90.6%; Pred. No. 4.7e-51;
Matches 252; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 22 TGACAGAAATTCACATTCGAGATGGCCCTCTAAATCCTGGCTGAATTTTAACTTCC 81
DB 40 TTACAGAAATTCACATTCGCGAAATGGCCCTCTAAATCCTGGCTGAATTTTAACTTCC 99
QY 82 TCTATGATCGGCAATAGGTTTATTTTATTTCTCAGCTACTAGTATTTTCTGGGAG 141
DB 100 TCTGTGATCAGCAATAGGATTTCTTTATGTTCTCAGCTATTAGTATTTCTGGGAG 159
QY 142 AAGAGGTGACACCCAGACTAATGTTCTTCAATGATCCTCATGCGAGGCATTCAGATG 201
```

```
BASE COUNT      624 a      300 c      334 g      536 t
ORIGIN
Query Match      79.1%; Score 236.4; DB 9; Length 1794;
Best Local Similarity 90.6%; Pred. No. 4.7e-51;
Matches 252; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
Qy 22 TGACAGAAATTCACCTTCCGAGATGGCCCTTAATCCTCGCGAATTTTAAACCTCC 81
Db 1111111111111111111111111111111111111111111111111111111
40 TTACAGAAATACACTTTTCGGGAAATGGCCCTAAATCCTGGCTGAATTTTAAACCTCC 99
Qy 82 TCATGATCGGCAATAGGCTTATTTTATTTCTCAGCTACTTAGTATTGTTGGGAG 141
Db 1111111111111111111111111111111111111111111111111111111
100 TCCTGATGATGAGCAATAGGATTTCTTTTATGTTCTCAGCTATTAGTATTGTTGGGAG 159
Qy 142 AAGAGGATGACACCCAGCACTAAATGTTCTTCAATATGATCCTCATGCCAGGCATTCAGATG 201
Db 1111111111111111111111111111111111111111111111111111111
160 AAGAGGTTGACACCCAGCCCTAATGTTCTTCAATATGATCCTCATGCCAGGCATTCAGATG 219
Qy 202 ATATGACAGATCATCTCTAGGAGGACAAATGAACCTTCAATGATGATCTAGCCAAAGCTA 261
Db 1111111111111111111111111111111111111111111111111111111
220 ATATGACAGATCATCTCTAGGAGGACAAATGAACCTTCAATGATGATCTAGCCAAAGCTA 279
Qy 262 AAGATGAGACACAGAAATCGCTCAAAACCTCTATTAG 299
Db 1111111111111111111111111111111111111111111111111111111
280 AAGATGAGACACAGACATGCTGCAAAACCTCTATCAG 317
Db 1111111111111111111111111111111111111111111111111111111

RESULT 8
HSAL32443      930 bp      mRNA      linear      PRI 27-JUL-2001
LOCUS
DEFINITION Homo sapiens mRNA for beta-1,3-galactosyltransferase b3Gal-T8.
ACCESSION AJ132443
VERSION AJ132443.1 GI:15028815
KEYWORDS b3Gal-T8; b3Gal-T8 gene; beta-1,3-galactosyltransferase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 930)
AUTHORS Anado,M., Almeida,R., Schwientek,T. and Clausen,H.
TITLE Identification and characterization of large galactosyltransferase
gene families: galactosyltransferases for all functions
JOURNAL Biochim. Biophys. Acta 1473 (1), 35-53 (1999)
MEDLINE 20047730
PUBMED 10580128
REFERENCE 2 (bases 1 to 930)
AUTHORS Jensen,M.A. and Bennett,E.P.
TITLE Cloning of a new member of the beta-1,3-galactosyltransferase
family, b3Gal-T8
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 930)
AUTHORS Jensen,M.A.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1999) Jensen M.A., School of Dentistry,
University of Copenhagen, Nørre Alle 20, 2200 Copenhagen N, DENMARK
FEATURES
Location/Qualifiers
1..930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7p14-p13"
1..930
/gene="b3Gal-T8"
1..930
/gene="b3Gal-T8"
/codon_start=1
/product="beta-1,3-galactosyltransferase b3Gal-T8"
/protein_id="CAC45046.1"
/db_xref="GI:15028816"
/translators="MASKSWLFLTLFCGSAIGLLCSOLFSLICEKVDTPNVLHN
DPHARSDDDNGQHLSEGQMFNADSSQHKDNDIAENLYQKVRILCWMTGQNLK
KAKHVRATQRCNKVLFMSSSEENKDFPVLGLTKRQDOLYKWTIKAFQYVHEHYLE
DADWFLKADDDTYVILDNLRLWLSKYDPEEPIYFGRFRKPYVRQGYMSGGAGYVLSKE"

BASE COUNT      312 a      161 c      192 g      265 t
ORIGIN
Query Match      77.3%; Score 231; DB 9; Length 930;
Best Local Similarity 94.1%; Pred. No. 1.2e-49;
Matches 240; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 45 ATGGCCCTCTAAATCCTCGCTGAATTTTAAACCTTCTCTATGATCGGCAATAGGCTTT 104
Db 1111111111111111111111111111111111111111111111111111111
1 ATGGCCCTCTAAATCCTCGCTGAATTTTAAACCTTCTCTCTGATCAGCAATAGGATTT 60
Qy 105 ATTTTATTTTCTCAGCTACTTAGTATTGTTGGGAGAAGAGGTTGACACCCAGCACTAAT 164
Db 1111111111111111111111111111111111111111111111111111111
61 CTTTATGTTCTCAGCTATTAGTATTGTTGGGAGAAAAGTTGACACCCAGCCCTAAT 120
Qy 165 GTTCTTCATATGATCCTCATGCCAGGCATTCAGATGATATGGACAGATCATCTAGGA 224
Db 1111111111111111111111111111111111111111111111111111111
121 GTTCTTCATATGATCCTCATGCCAGGCATTCAGATGATATGGACAGATCATCTAGAA 180
Qy 225 GGACAAATGAACCTTCAATGATGATCTAGGCCAAGCTTAAAGATGAGAACACAGACATTCCT 284
Db 1111111111111111111111111111111111111111111111111111111
181 GGACAAATGAACCTTCAATGATGATCTAGGCCAAGCTTAAAGATGAGAACACAGACATTCCT 240
Qy 285 GAAACCTCTATTAG 299
Db 1111111111111111111111111111111111111111111111111111111
241 GAAACCTCTATCAG 255
Db 1111111111111111111111111111111111111111111111111111111

RESULT 9
AXI175170      1092 bp      DNA      linear      PAT 03-JUL-2001
LOCUS
DEFINITION Sequence 12 from Patent WO0144478.
ACCESSION AXI175170
VERSION AXI175170.1 GI:14598568
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1092)
AUTHORS Canfield,W.M., Cummings,R.D. and Ju,T.
TITLE Core lb3-galactosyl transferases and methods of use thereof
JOURNAL Patent: WO 0144478-A 12 21-JUN-2001;
The Board of Regents of The University of Oklahoma (US)
FEATURES
Location/Qualifiers
1..1092
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT      363 a      195 c      219 g      315 t
ORIGIN
Query Match      77.3%; Score 231; DB 6; Length 1092;
Best Local Similarity 94.1%; Pred. No. 1.2e-49;
Matches 240; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 45 ATGGCCCTCTAAATCCTCGCTGAATTTTAAACCTTCTCTATGATCGGCAATAGGCTTT 104
Db 1111111111111111111111111111111111111111111111111111111
1 ATGGCCCTCTAAATCCTCGCTGAATTTTAAACCTTCTCTCTGATCAGCAATAGGATTT 60
Qy 105 ATTTTATTTTCTCAGCTACTTAGTATTGTTGGGAGAAGAGGTTGACACCCAGCACTAAT 164
Db 1111111111111111111111111111111111111111111111111111111
61 CTTTATGTTCTCAGCTATTAGTATTGTTGGGAGAAAAGTTGACACCCAGCCCTAAT 120
Qy 165 GTTCTTCATATGATCCTCATGCCAGGCATTCAGATGATATGGACAGATCATCTAGGA 224
Db 1111111111111111111111111111111111111111111111111111111
121 GTTCTTCATATGATCCTCATGCCAGGCATTCAGATGATATGGACAGATCATCTAGAA 180
Qy 225 GGACAAATGAACCTTCAATGATGATCTAGGCCAAGCTTAAAGATGAGAACACAGACATTCCT 284
Db 1111111111111111111111111111111111111111111111111111111
181 GGACAAATGAACCTTCAATGATGATCTAGGCCAAGCTTAAAGATGAGAACACAGACATTCCT 240
```

```

Qy 285 GAAACCTCTATTAG 299
|||||
Db 241 GAAACCTCTATCAG 255

RESULT 10
LOCUS HSA243256 1092 bp mRNA linear PRI 04-JAN-2002
DEFINITION Homo sapiens mRNA for beta-1,3-galactosyltransferase.
ACCESSION AJ243256
VERSION AJ243256.1 GI:18073134
KEYWORDS b3Gal-T gene; beta-1,3-galactosyltransferase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Jensen, M.P.A.
TITLE Cloning and expression of a novel beta-1,3-Galactosyltransferase
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1092)
AUTHORS Jensen, M.P.A.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-1999) Jensen M.P.A., School of Dentistry,
University of Copenhagen, Noerre Alle 20, DK-2200 Copenhagen N,
DENMARK
FEATURES
Source Location/Qualifiers
1..1092
/organism="Homo sapiens"
/db_xref="taxon:9606"
/gene="b3Gal-T"
/feature="making the T-antigen"
/codon_start=1
/product="beta-1,3-Galactosyltransferase"
/protein_id="CAC80435.1"
/db_xref="GI:18073135"
/translation="MASKSWLNFLTFLCGSAIGLLCSQLFSILLGKVDTPQPNVLHN
DPHARHSDNGQNHLEQOMFNADSSQHKDENTDIAENLYQKVRILCWMTGPNLEK
KAKHVATQAQRCKNVLFMSSEENKDPFVGLKTKGRDOLYKRTIKAFQYVHEHYLE
DAMFLKADDDTYVLDNLRLWLLSKYDPEEPIYFGRFPKPYKGYMSGGAGYVLSKE
ALKRFVDAFTDKTCHSSSIDLALGRCEIMNVAGDSRDTIGKETFHFPVPEHHLI
KGYLPRTFWNTNYPVPVGGCCSLAVSFHYVDSFTTMYELEYLVHLRPGYLYR
QYPLPRILKEISQANKNEDTKVLGNP"
BASE COUNT 363 a 195 c 219 g 315 t
ORIGIN
Query Match 77.3%; Score 231; DB 9; Length 1092;
Best Local Similarity 94.1%; Pred. No. 1.2e-49;
Matches 240; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 45 ATGCCCTCTAAATCCTCGCTGAATTTTAAACCTTCCTCTATGATCGGCAATAGGCTTT 104
|||||
Db 1 ATGCCCTCTAAATCCTCGCTGAATTTTAAACCTTCCTCTATGATCGGCAATAGGCTTT 60
|||||
Qy 105 ATTTTATTTCTCAGCTACTAGTATTCTTGTGGAGAGAGGTGACACCCAGCTAAT 164
|||||
Db 61 CTTTATGTCTCAGCTATTAGTATTCTTGTGGAGAGAGGTGACACCCAGCTAAT 120
|||||
Qy 165 GTTCTTCATATGATCTCTATGCGAGCATTCAGATGATATGACAGAAATCATCTAGGA 224
|||||
Db 121 GTTCTTCATATGATCTCTATGCGAGCATTCAGATGATATGACAGAAATCATCTAGGA 180
|||||
Qy 225 GGACAAATGAACCTTCAATTCAGATTTCTAGCCAACTGTAAGATGAGAACACAGAAATCGCT 284
|||||
Db 181 GGACAAATGAACCTTCAATTCAGATTTCTAGCCAACTGTAAGATGAGAACACAGAAATCGCT 240
|||||
Qy 285 GAAACCTCTATTAG 299
|||||
Db 241 GAAACCTCTATCAG 255

RESULT 11
LOCUS HSA278960 1092 bp mRNA linear PRI 05-DEC-2001
DEFINITION Homo sapiens mRNA for beta 1,3-galactosyltransferase (core-1-beta
1,3-galactosyltransferase gene).
ACCESSION AJ278960
VERSION AJ278960.1 GI:17384685
KEYWORDS beta 1,3-galactosyltransferase; core-1-beta
1,3-galactosyltransferase gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Bennett, E.P.
TITLE Unpublished
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1092)
AUTHORS Bennett, E.P.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-2000) Bennett E.P., Histo.lab., 24.5.33, School
of Dentistry, Noerre Alle' 20, 2200N, Copenhagen, DENMARK
FEATURES
Source Location/Qualifiers
1..1092
/organism="Homo sapiens"
/db_xref="taxon:9606"
/gene="core-1-beta 1,3-galactosyltransferase"
/feature="Core-1 synthase"
/codon_start=1
/evidence="experimental"
/product="beta 1,3-galactosyltransferase"
/protein_id="CAC82373.1"
/db_xref="GI:17384686"
/translation="MASKSKWLNFLTFLCGSAIGLLCSQLFSILLGKVDTPQPNVLHN
DPHARHSDNGQNHLEQOMFNADSSQHKDENTDIAENLYQKVRILCWMTGPNLEK
KAKHVATQAQRCKNVLFMSSEENKDPFVGLKTKGRDOLYKRTIKAFQYVHEHYLE
DAMFLKADDDTYVLDNLRLWLLSKYDPEEPIYFGRFPKPYKGYMSGGAGYVLSKE
ALKRFVDAFTDKTCHSSSIDLALGRCEIMNVAGDSRDTIGKETFHFPVPEHHLI
KGYLPRTFWNTNYPVPVGGCCSLAVSFHYVDSFTTMYELEYLVHLRPGYLYR
QYPLPRILKEISQANKNEDTKVLGNP"
BASE COUNT 363 a 195 c 219 g 315 t
ORIGIN
Query Match 77.3%; Score 231; DB 9; Length 1092;
Best Local Similarity 94.1%; Pred. No. 1.2e-49;
Matches 240; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 45 ATGCCCTCTAAATCCTCGCTGAATTTTAAACCTTCCTCTATGATCGGCAATAGGCTTT 104
|||||
Db 1 ATGCCCTCTAAATCCTCGCTGAATTTTAAACCTTCCTCTATGATCGGCAATAGGCTTT 60
|||||
Qy 105 ATTTTATTTCTCAGCTACTAGTATTCTTGTGGAGAGAGGTGACACCCAGCTAAT 164
|||||
Db 61 CTTTATGTCTCAGCTATTAGTATTCTTGTGGAGAGAGGTGACACCCAGCTAAT 120
|||||
Qy 165 GTTCTTCATATGATCTCTATGCGAGCATTCAGATGATATGACAGAAATCATCTAGGA 224
|||||
Db 121 GTTCTTCATATGATCTCTATGCGAGCATTCAGATGATATGACAGAAATCATCTAGGA 180
|||||
Qy 225 GGACAAATGAACCTTCAATTCAGATTTCTAGCCAACTGTAAGATGAGAACACAGAAATCGCT 284
|||||
Db 181 GGACAAATGAACCTTCAATTCAGATTTCTAGCCAACTGTAAGATGAGAACACAGAAATCGCT 240
|||||
Qy 285 GAAACCTCTATTAG 299
|||||
Db 241 GAAACCTCTATCAG 255

```


RESULT 12

AC005532
 LOCUS Homo sapiens PAC clone RP4-733B9 from 7p14-p13, complete sequence.
 DEFINITION AC005532
 ACCESSION AC005532
 VERSION AC005532.2 GI:21322201
 KEYWORDS HTG.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 115954)
 Sulston, J.E. and Waterston, R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

99063792

PUBMED

9847074

2 (bases 1 to 115954)

Dauphin, S. and Gregory, S.

The sequence of Homo sapiens PAC clone RP4-733B9

Unpublished (2001)

3 (bases 1 to 115954)

Waterston, R.H.

Direct Submission

Submitted (25-AUG-1998) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

4 (bases 1 to 115954)

Waterston, R.

Direct Submission

Submitted (14-JAN-1999) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 5 (bases 1 to 115954)

Waterston, R.

Direct Submission

Submitted (21-DEC-1999) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 6 (bases 1 to 115954)

Waterston, R.

Direct Submission

Submitted (04-JUN-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Jun 4, 2002 this sequence version replaced gi:4156192.

COMMENT

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@watson.wustl.edu

----- Summary Statistics

----- Center project name: H_DJ0733B09

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:

The sequencing of this clone was established as part of a mapping and
 sequencing collaboration between the NHGRI Chromosome 7 Mapping
 Project (Eric D. Green, Director), John D. McPherson in the
 Department of Genetics (Washington University), and the Washington
 University Genome Sequencing Center. For additional information
 about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send

<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by
 Pieter de Jong and coworkers at <http://www.chori.org> using the
 method described by Ioannou et al., Nature Genetics 6:84-9 (1994).
 The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc.

(<http://www.genomesystems.com>) or Research Genetics, Inc.

(<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The actual start of this clone is at base position 1 of RP4-733B9,
 actual end is at 115954 of RP4-733B9.

Base positions 26118 to 26479 and 25639 to 25745 of RP4-733B9 are
 covered by per product from RP4-733B9 clone DNA.

FEATURES

Source	Location/Qualifiers
1..115954	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7p14-p13"
	/clone="RP4-733B9"
	/clone_lib="RPCI-4"
repeat_region	1..221
	/rpt_family="L1"
repeat_region	222..520
	/rpt_family="Alu"
repeat_region	521..2344
	/rpt_family="L1"
repeat_region	2364..2774
	/rpt_family="L1"
repeat_region	2775..3195
	/rpt_family="MaLR"
repeat_region	3196..3297
	/rpt_family="L1"
repeat_region	3298..3342
	/rpt_family="MaLR"
repeat_region	3344..3630
	/rpt_family="Alu"
repeat_region	3646..3993
	/rpt_family="MaLR"
repeat_region	3998..4184
	/rpt_family="L1"
repeat_region	4185..4316
	/rpt_family="Alu"
repeat_region	4317..4361
	/rpt_family="L1"
repeat_region	4357..5291
	/rpt_family="L1"
repeat_region	5392..5686
	/rpt_family="MaLR"
repeat_region	5671..5890
	/rpt_family="MER21-group"
repeat_region	7540..7699
	/rpt_family="MIR"
repeat_region	8200..8227
	/rpt_family="AT-rich"
repeat_region	8458..8518
	/rpt_family="L2"
repeat_region	8467..8525
	/rpt_family="MIR"
repeat_region	8531..9081
	/rpt_family="MaLR"
repeat_region	9397..9829
	/rpt_family="MaLR"
repeat_region	10026..10197
	/rpt_family="L1"
repeat_region	10465..10645
	/rpt_family="Alu"
repeat_region	10791..10822

Matches	218;	Conservative	0;	Mismatches	40;	Indels	0;	Gaps	0;
QY	42	GAGATGCCCTCTAAATCCTGGCTGAATTTTTAAACCTTCTCTATGGATCGGCAATAGGG	101						
Db	146	GAAATGCCCTCTAAATCTTGGCTGAATTTTTTAACTCTCTCTGGATCAGCAATAGGG	205						
QY	102	TTTATTTTATTTTCTCAGCTACTAGTATTTCTTGGGAGAAGAGGGTGACACCCAGACT	161						
Db	206	TTTTTTTTTATGTTCTCAACTCTTGAGTATTTGTGGGAGAAGAGGCTGCCATTCAGCCT	265						
QY	162	AATGTTCTTCATATGATCCTCATCGGAGCATTCAGATGATATATGGACAGAATCATCTA	221						
Db	266	AACATGCTTCACAAATGACCCCTCATGCAAGGCATTCAGATGCAATGACACAGTCACCTC	325						
QY	222	GGAGGACAATGAACCTCAATGCAGATCTTAGCCCAAGTAAAGATGAGACACAGAAATC	281						
Db	326	AAGGACAGATGAACCTTCATGCAGATTCAGCCACATAAAGATGAGAATAGAGCTT	385						
QY	282	GCTGAAAACCTCTATTAG	299						
Db	386	GCTGAGAACCCTCTATCAG	403						

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 22:29:21 ; Search time 117.627 Seconds
(without alignments)
5724.434 Million cell updates/sec

Title: US-10-029-359A-1

Perfect score: 299

Sequence: 1 gcaagaggagccacggcg.....tcgtgaaacacctattag 299

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*			
1:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*		
2:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*		
3:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*		
4:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*		
5:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*		
6:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*		
7:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*		
8:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*		
9:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*		
10:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*		
11:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*		
12:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*		
13:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*		
14:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*		
15:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*		
16:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*		
17:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*		
18:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*		
19:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*		
20:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*		
21:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*		
22:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*		
23:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*		
24:	/SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	298	99.7	505 24 ABK88407 Human DNA encoding
2	236.4	79.1	1794 22 AAF90507 Human core 1 beta3
3	236.4	79.1	1800 21 AAZ99551 Nucleotide sequenc
4	224.2	75.0	1424 23 AAS73202 DNA encoding novel
5	194	64.9	1469 22 AAF90509 Mouse core 1 beta3
6	189.2	63.3	1440 22 AAF90508 Rat core 1 beta3-g
7	91	30.4	1560 24 ABK88409 Human DNA encoding
8	66.2	22.1	795 22 AAL24100 Human breast cance
9	46.2	15.5	489 24 ABK88408 Human DNA encoding

10	37.6	12.6	2646	16	AAT02405	Rice mature pullul
11	37.6	12.6	2982	16	AAT02400	Pullulanase expres
12	37.6	12.6	2988	16	AAT02399	Rice mature pullul
13	36	12.0	4282	21	AAC48551	Arabidopsis thalia
14	35.6	11.9	10859	22	AAL04720	Human reproductive
15	35.6	11.9	10859	23	ABL97627	Human testicular a
16	35.4	11.8	1083	23	AAL76745	DNA encoding novel
17	35.2	11.8	513	22	AAL15254	Human breast, cance
18	34.2	11.4	4265	21	AAH29796	S cerevisiae apopt
19	34	11.4	342	22	AAA44815	Human secreted exp
20	34	11.4	4764	22	ABA08912	Human cancer assoc
21	34	11.4	4764	22	AAH90024	Human bone marrow
22	34	11.4	4767	22	AAH89911	Human bone marrow
23	34	11.4	5807	21	AAA09157	Human cancer assoc
24	34	11.4	6380	21	AAA09159	Human cancer assoc
25	34	11.4	11314	23	ABL06868	Drosophila melanog
26	34	11.4	17666	22	AKK76302	Human immune PCR-
27	33.8	11.3	618	22	AAS23915	Human ovarian tumo
28	33.8	11.3	618	22	AAH82472	DNA encoding novel
29	33.8	11.3	1782	23	AAS78516	Human cDNA for nov
30	33.8	11.3	2001	24	ABK34618	DNA encoding novel
31	33.8	11.3	2263	23	AAS92449	DNA encoding novel
32	33.8	11.3	18692	23	ABK42795	Genomic sequence #
33	33.6	11.2	3647	23	ABL18680	Drosophila melanog
34	33.6	11.2	4153	23	ABL12654	Drosophila melanog
35	33.4	11.2	6935	24	ABL92262	Chemically treated
36	33.4	11.2	47066	23	ABL11514	Drosophila melanog
37	33.4	11.2	34980	22	AAH41225	Pyrococcus abyssi
38	33.2	11.1	2123	20	AAH20304	Borrelia burgdorfe
39	33.2	11.1	26241	22	ABA16222	Human nervous syst
40	33.2	11.1	147708	24	ABQ88154	Human osteoblast d
41	33	11.0	837	22	AAH53653	S. epidermidis ope
42	33	11.0	3236	22	AAH54128	S. epidermidis gen
43	33	11.0	9145	24	AAH32888	Human immune syste
44	32.8	11.0	1707	22	AAH45215	Human zinc finger
45	32.8	11.0	1919	22	AAK03010	Human brain expres

ALIGNMENTS

RESULT 1

ABK88407

ID ABK88407 standard: DNA; 505 BP.

XX AC ABK88407;

XX DF 07-OCT-2002 (first entry)

XX DE Human DNA encoding partial protein cluster II protein #1.

XX KW Human; ds: gene; protein cluster II; obesity; diabetes mellitus;

XX KW central nervous system disorder; metabolic disease.

XX OS Homo sapiens.

XX FH Key

XX CDS Location/Qualifiers

FT CDS 21..497

FT /*tag= a

FT /product= "protein cluster II protein #1"

FT /note= "No stop codon shown"

FT /transl_except= (pos:297..299,aa:aa)

FT /note= "Xaa= unknown"

XX PN WO200251864-A1.

XX PD 04-JUL-2002.

XX PF 14-DEC-2001; 2001WO-SE02786.

XX PR 22-DEC-2000; 2000SE-0004828.

XX PA (PHAA) PHARMACIA AB.

XX PI Attersand A;
XX DR WPI: 2002-575368/61.
XX DR P-PSDB; ABG30851.
XX PT New protein cluster II nucleic acids and polypeptides, useful in
XX PT diagnosing metabolic diseases such as obesity and diabetes, and in
XX PT identifying agents for treating such diseases
XX PS Claim 1; Page 20-21; 28pp; English.
XX CC The invention relates to a new isolated nucleic acid comprising:
XX CC (a) a fully defined sequence, encoding a protein cluster II protein,
XX CC appearing as ABK8407 and ABK8408; (b) a sequence capable of hybridising
XX CC under stringent hybridisation conditions to a nucleotide sequence
XX CC complementary to the polypeptide coding region of a nucleic acid in (a);
XX CC or (c) a sequence which is degenerate as a result of the genetic code to
XX CC a nucleotide sequence in (a) or (b). Also included are a isolated
XX CC an polypeptide encoded by the nucleic acid, a vector harbouring the
XX CC nucleic acid, a replicable expression vector, which carries and is
XX CC capable of mediating the expression of the nucleic acid a cultured host
XX CC cell harbouring the expression vector, producing a polypeptide by
XX CC culturing the cell, where the polypeptide is produced, and recovering the
XX CC polypeptide and identifying an agent capable of modulating the nucleic
XX CC acid by providing a cell comprising the nucleic acid molecule, contacting
XX CC the cell with a candidate agent, and monitoring the cell for an effect
XX CC that is not present in the absence of the candidate agent. The protein
XX CC cluster II nucleic acid and polypeptide is useful in the diagnosis of
XX CC metabolic diseases such as obesity and diabetes, and central nervous
XX CC system disorders and in the identification of agents for treating these
XX CC diseases. The nucleic acids may be used as hybridisation probes, for
XX CC chromosome and gene mapping, in polymerase chain reaction (PCR)
XX CC technologies, in the production of sense and antisense nucleic acids, and
XX CC in screening for new therapeutic molecules. The present sequence
XX CC encodes a protein cluster II protein of the invention.
XX SQ Sequence 505 BP; 166 A; 95 C; 108 G; 135 T; 1 other;
Query Match 99.7%; Score 298; DB 24; Length 505;
Best Local Similarity 99.7%; Pred. No. 9.4e-81;
Matches 298; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCAAGAGGAGCGACGGCGATGACAGAAAATTCACCTTCCGAGATGCCCTTAATCCT 60
Db 1 GCAAGAGGAGCGACGGCGATGACAGAAAATTCACCTTCCGAGATGCCCTTAATCCT 60
QY 61 GGCTCAATTTTAACTTCTCTATGATCGGCAATAGGGTTATTTTATTTCTCAGC 120
Db 61 GGCTCAATTTTAACTTCTCTATGATCGGCAATAGGGTTATTTTATTTCTCAGC 120
QY 121 TACTTAGTATTTTGGGAGAGAGGGTGACACCCAGACTAATCTTCTCATATGATC 180
Db 121 TACTTAGTATTTTGGGAGAGAGGGTGACACCCAGACTAATCTTCTCATATGATC 180
QY 181 CTCATGCGAGGATTCAGATGATATGACAGAAATCATCTAGGAGGACAAATGAACCTCA 240
Db 181 CTCATGCGAGGATTCAGATGATATGACAGAAATCATCTAGGAGGACAAATGAACCTCA 240
QY 241 ATGCAGATTCAGCAACGTAAGATGAGACACAGAAATCGCTCAAAACCTCTATTAG 299
Db 241 ATGCAGATTCAGCAACGTAAGATGAGACACAGAAATCGCTCAAAACCTCTATTAG 299
RESULT 2
ID AAF90507
XX AAF90507 standard; cDNA; 1794 BP.
XX AC AAF90507;
XX XX
XX XX 22-AUG-2001 (first entry)
XX DE Human core 1 beta3-galactosyl transferase cDNA.

XX KW Core 1 beta3-galactosyl transferase; human; O-glycosylation;
XX KW galactosylation; glycosulfopeptide; Tn syndrome; IgA nephropathy;
XX KW diagnosis; therapy; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT 63..1154
XX FT /*tag= a
XX FT /note= "a polynucleotide comprising this
XX FT coding sequence is also claimed in Claim
XX FT 1(A)"
XX PN WO200144478-A2.
XX PD 21-JUN-2001.
XX PF 14-DEC-2000; 2000WO-US33945.
XX PR 15-DEC-1999; 99US-0461321.
XX PR 15-DEC-1999; 99US-0464035.
XX PA (UYOK-) UNIV OKLAHOMA.
XX PI Canfield WM, Cummings RD, Ju T;
XX DR WPI: 2001-398157/42.
XX DR P-PSDB; AAB82456.
XX CC Novel purified core 1 beta3 galactosyl transferases of human, rat,
XX CC mouse, Drosophila melanogaster, Caenorhabditis elegans, useful for
XX CC galactosylating N-acetyl galactosamine linked to O-linking amino acid
XX CC on protein
XX PS Claim 1(A); Fig 3; 74pp; English.
XX CC The present sequence is that of human cDNA for core 1
XX CC beta3-galactosyltransferase (b3GTase, see AAB82456). The coding
XX CC region of the sequence is also claimed. The cDNA was identified
XX CC initially as an expressed sequence tag (EST) following BlastN
XX CC searching with a rat EST sequence, and was completed by 5'-RACE
XX CC using human placenta cDNA as template. The encoded protein is a
XX CC type 2 transmembrane protein. The invention provides human, rat,
XX CC mouse, Drosophila melanogaster and Caenorhabditis elegans b3GTases
XX CC and nucleic acids encoding them. The polypeptides exhibit a wide
XX CC range of homologies. The polynucleotides can be used to transform
XX CC or transfect host cells for producing substantially pure forms of
XX CC the enzyme, or for use in an expression system for core 1 O-linked
XX CC glycosylation of proteins or peptides produced within the
XX CC expression system. The expressed enzymes galactosylate, via a
XX CC beta3 linkage, an N-acetyl galactosamine linked to a serine,
XX CC threonine or other O-linking amino acid on peptides or proteins.
XX CC Core 1 b3GTases are useful in the synthesis of glycosulfopeptides
XX CC which can function as inhibitors of P-selectin:PSGL-1 interactions.
XX CC Other potential uses include diagnostic tests for the rare
XX CC Tn-syndrome or IgA nephropathy, and the therapy of these disorders.
XX SQ Sequence 1794 BP; 624 A; 300 C; 334 G; 536 T; 0 other;
Query Match 79.1%; Score 236.4; DB 22; Length 1794;
Best Local Similarity 90.6%; Pred. No. 8.3e-62;
Matches 252; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 22 TCACAGAAATTCACCTTCCGAGATGGCTCTAATCTCGCTGAATTTTAACTTCC 81
Db 40 TTACAGAAATACACTTTCGGGAATGGCTCTAATCTCGCTGAATTTTAACTTCC 99
QY 82 TCTATGGATCGCAATAGGGTTTATTTATTTTCTCAGCTACTAGTATTTGTGGGAG 141
Db 100 TCTGTGGATCAGCAATAGGATTTCTTTATGTCTCAGCTATTTAGTATTTGTGGGAG 159
QY 142 AAGAGGGTGACACCCAGACTAATGTCTTCATATGATCTCATCGGAGGCAATTCAGATG 201

160 AAAAGGTTGACACCCAGCCTAATGTTCTCATAATGATCTCTGCAAGCATTCAGATG 219
202 ATAATGACAGAAATCATCTAGGAGGACAAATGAATCAATGAGATTCATGACCAACGTA 261
220 ATAATGACAGAAATCATCTAGGAGGACAAATGAATCAATGAGATTCATGACCAACATA 279
262 AAGATGAGACACAGAAATCGCTGAAACCTCTATTAG 299
280 AAGATGAGACACAGAAATGCTGTAACCTCTATCAG 317

RESULT 3

AAZ99551
ID AAZ99551 standard; DNA; 1800 BP.

AAZ99551;

03-JUL-2000 (first entry)

Nucleotide sequence of human core-1 2 beta1,3-GalT.

Glycosulfopeptide; P-selectin glycoprotein ligand-1;
inflammatory disease; chronic inflammation; acute inflammation;
diffuse inflammation; traumatic inflammation; immunosuppression;
toxic diffuse inflammation; specific inflammation; core-1 2 beta1,3-GalT;
reactive inflammation; parenchymatous inflammation;
obliterative inflammation; interstitial inflammation;
croupous inflammation; focal inflammation; rheumatoid arthritis;
acute leukocyte-mediated lung injury; acute inflammation; ss.

Homo sapiens.

Key Location/Qualifiers

63..1154

/*tag= a

/product= "core-1 2 beta1,3-GalT"

WO9965712-A2.

23-DEC-1999.

15-JUN-1999; 99WO-US13455.

16-JUN-1998; 98US-0089472.

(CUMM//) CUMMINGS R D.
(MCEV//) MCEVER R P.

Cummings RD, McEver RP;

WPI; 2000-237146/20.

P-PSDB; AAY84165.

Novel synthetic glycosulfopeptides which are P- selectin glycoprotein
ligand-1 (PSGL-1) mimics, used as anti- inflammatory agents in the
treatment of acute and chronic inflammation

Example 1; Page 73-75; 80pp; English.

The present sequence encodes a human core-1 2 beta1,3-GalT polypeptide.
The polypeptide is used to produce the synthetic glycosulfopeptides of
the invention. The specification describes a new class of synthetic
glycosulfopeptides which mimic the extreme amino terminus of P-selectin
glycoprotein ligand-1. The glycosulfopeptides comprise one or more
sulphated tyrosine residues and a glycan comprising a sialyl Lewis-x
group or a sialyl Lewis-a group. The synthetic glycosulfopeptides are
used to treat inflammatory diseases. They can be used to treat both
chronic and acute inflammatory diseases. e.g. diffuse inflammation, traumatic
inflammation, immunosuppression, toxic diffuse inflammation, traumatic
inflammation, reactive inflammation, parenchymatous inflammation, specific
obliterative inflammation, interstitial inflammation, croupous
inflammation and focal inflammation. Diseases which may be treated

CC include rheumatoid arthritis, post-ischemic (reperfusion)
CC leukocyte-mediated tissue damage, acute leukocyte-mediated lung
CC injury (e.g. Adult respiratory Distress syndrome), and other tissue or
CC organ specific forms of acute inflammation (e.g. glomerulonephritis).
CC The GSPs can also be used with enzyme linked immunosorbant assay (ELISA)
CC techniques to distinguish between monoclonal antibodies which react with
CC core-2 sialyl Lewis-x groups versus those which react with core-1 sialyl
CC Lewis-x groups. The glycosulfopeptides are also excellent acceptors for
CC specific glycosyltransferases.

XX Sequence 1800 BP; 625 A; 302 C; 336 G; 537 T; 0 other;

Query Match 79.1%; Score 236.4; DB 21; Length 1800;

Best Local Similarity 90.6%; Pred. No. 8.3e-62;

Matches 252; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 22 TGACAGAAAATTCACATTTCCGAGATGGCTCTAAATCCTGGCTGAATTTTAACTTCC 81

Db 40 TTACAGAAATACACTTTTCGGAAATGGCTCTAAATCCTGGCTGAATTTTAACTTCC 99

QY 82 TCTATGGATCGGCAATAGGGTTTATTTATTTCTCAGCTACTTAGTATTTTGGGAG 141

Db 100 TCTGTGGATCAGCAATAGGATTTCTTTATCTCTCAGCTATTTAGTATTTTGGGAG 159

QY 142 AAGAGGGTGACACCCAGACTAATGTTCTTATATGATCTCTCAGGAGGATTCAGATG 201

Db 160 AAAAGGTTGACACCCAGCGCTAATGTTCTTCTATAATGATCTCTCAGGAGGATTCAGATG 219

QY 202 ATAATGGACAGAATCATCTAGGAGCACAATAATGAATTCATGAGATCTAGCCAAAGTA 261

Db 220 ATAATGGACAGAATCATCTAGGAGCACAATAATGAATTCATGAGATCTAGCCAAAGTA 279

QY 262 AAGATGAGACACAGAAATCGCTGAAAACCTCTATTAG 299

Db 280 AAGATGAGACACAGAAATCGCTGAAAACCTCTATTAG 317

RESULT 4

AAZ93202

ID AAZ93202 standard; cDNA; 1424 BP.

XX AC AAZ93202;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #9006.

Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder; ss.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

P-PSDB; ABG09015.

New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX

Claim 1; SEQ ID No 9006; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS45464 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 1424 BP; 456 A; 285 C; 315 G; 368 T; 0 other;

Query Match	75.08;	Score 224.2;	DB 23;	Length 1424;
Best Local Similarity	92.9%;	Pred. No. 3.9e-58;		
Matches 235;	Conservative 0;	Mismatches 18;	Indels 0;	Gaps 0;
QY	37	TTTCCGAGATGGCCTCTAAATCCFTGGCTGAATTTTTTAAACCTTCCTATGCGATCGCAA	96	
Db	9	TTCCGGAAATGGCCTCTAAATCCTGGCTGAATTTTTTAACTTCCTCTGTGGATCAGCAA	68	
QY	97	TAGGTTTATTTATTTTCTCAGCTACTTAGTATTTTGTGGGAGAGAGGGTGACACCC	156	
Db	69	TAGGATTTCTTTATGTCTCAGCTATTTAGTATTTTGTGGGAGAAAAGGTTGACACCC	128	
QY	157	AGACTAATGTCTTCATAATGATCCTCATCGAGGCATTCAGATGATAATGCACAGAATC	216	
Db	129	AGCCTAATGTCTTCATAATGATCCTCATCGAGGCATTCAGATGATAATGGACAGAATC	188	
QY	217	ATCTAGGAGACAAATGAACCTTCAATCCAGATTCCTAGCCCAACGTAAAGATGAGAACACAG	276	
Db	189	ATCTAGGAAGACAAATGAACCTTCAATGCAGATTCCTAGCCCAACATAAAGATGAGAACACAG	248	
QY	277	AAATCGCTGAAAA	289	
Db	249	ACATTTGCTGAAAA	261	

RESULT 5
AAE90509
ID AAE90509 standard; cDNA: 1469 BP.

DT 22-AUG-2001 (first entry)

DE .Mouse core 1 beta3-galactosyl transferase cDNA.

Core 1 beta3-galactosyl transferase; mouse; O-glycosylation;
galactosylation; glycosulfopeptide; Tn syndrome; IgA nephropathy;
diagnosis; therapy; ss.

OS Mus musculus.

Key	Location/Qualifiers
XX	
FH	
FT	180..1271
FT	
FT	*tag= a
FT	/note= "a polynucleo
FT	coding sequen
FT	l(A)"
FT	

XX	WO200144478-A2.
PN	
XX	21-JUN-2001.
PD	
XX	14-DEC-2000; 200OWO-US33945.
PF	
XX	15-DEC-1999; 99US-0461321.
PR	
XX	15-DEC-1995; 99US-0464035.
PR	
XX	(UYOK-) UNTV OKLAHOMA.
PA	
XX	Canfield WM, Cummings RD, Ju T;
PI	
XX	WPI: 2001-398157/42.
DR	P-PSDB; AAB82458.
XX	
PT	Novel purified core 1 beta3 galactosyl transferases of human, rat,
PT	mouse, Drosophila melanogaster, Caenorhabditis elegans, useful for
PT	galactosylating N-acetyl galactosamine linked to O-linking amino acid
PT	on protein
XX	
PS	Claim 1(A): Page 64; 74pp; English.
XX	
CC	The present sequence is that of mouse cDNA for core 1
CC	beta3-galactosyltransferase (b3GTase, see AAB82458). The mouse
CC	core 1 b3GTase has 89% identity and 94% similarity to the human
CC	enzyme (see AAB82458). The invention provides human, rat, mouse,
CC	Drosophila melanogaster and Caenorhabditis elegans b3GTases and
CC	nucleic acids encoding them. The polypeptides exhibit a wide
CC	range of homologies. The polynucleotides can be used to transform
CC	or transfect host cells for producing substantially pure forms of
CC	the enzyme, or for use in an expression system for core 1 O-linked
CC	glycosylation of proteins or peptides produced within the
CC	expression system. The expressed enzymes can be used to
CC	galactosylate, via a beta3 linkage, an N-acetylgalactosamine linked
CC	to a serine, threonine or other O-linking amino acid on peptides or
CC	proteins. Core 1 b3GTases are also useful in the synthesis of
CC	glycosulfopeptides which can function as inhibitors of
CC	P-selectin:PSGL-1 interactions. Other potential uses include
CC	diagnostic tests for the rare fn-syndrome or Iga nephropathy, and
CC	therapy of these disorders.
XX	
SQ	Sequence 1469 BP; 467 A; 296 C; 313 G; 393 T; 0 other;
	Query Match 64.9%; Score 194; DB 22; Length 1469;
	Best Local Similarity 84.5%; Pred. No. 6.le-49;
	Matches 218; Conservative 0; Mismatches 40; Indels 0; Gaps 0
QY	42 GAGATGGCCTCTAAATCCTGGCTCAATTTTAAACCTTCCTCTATGATCGCAATAGG 101
Db	
	177 GAATGGCCTCTAAATCTTGCTGAATTTTTTAGCTCTCTCTGTCGATCAGCATAGG 236
QY	102 TTATTATTTATTTCTCAGCTACTTAGTATTTGTTGGGAAGAAGGGTGACACCAGACT 161
Db	
	237 TTTTFTTTTATCTTCTCAACTCTTGAGTATTTGTTGCGAAGAGSGCTGCCATTTCAGCCT 296
QY	162 ATGTTCTTCATATGATCCTCATGCAGGCATTCAGATGTAATGGACAGAAATCATCTA 221
Db	
	297 AACATGCTTCCAATGACCCCTCATGCAAGGCATTCAGATGACAATGGACACACAGTC 356
QY	222 GGAGGCAAAATGAACCTTCAATGCAGATTCTTAGCCACAGTAAAGATCAGAACACAGAAATC 281
Db	
	357 AAAGCAGATGAACCTTCATGTCAGATTCACGCCACACATAAAGATGAGAACATAGACGTT 416
QY	282 GCTGAAAACCTCTATTAG 299
Db	
	417 GCTGAGAACCCTCTATCAG 434
RESULT 6	
AAF90508	
AAF90508 standard; cdNA: 1440 BP.	

[illegible]

cell harbouring the expression vector, producing a polypeptide by culturing the cell, where the polypeptide is produced, and recovering the polypeptide and identifying an agent capable of modulating the nucleic acid by providing a cell comprising the nucleic acid molecule, contacting the cell with a candidate agent, and monitoring the cell for an effect that is not present in the absence of the candidate agent. The protein cluster II nucleic acid and polypeptide is useful in the diagnosis of metabolic diseases such as obesity and diabetes, and central nervous system disorders and in the identification of agents for treating these diseases. The nucleic acids may be used as hybridisation probes, for chromosome and gene mapping, in polymerase chain reaction (PCR) technologies, in the production of sense and antisense nucleic acids, and in screening for new therapeutic molecules. The present sequence encodes a protein cluster II protein of the invention.

Sequence 489 BP; 187 A; 86 C; 95 G; 121 T; 0 other;

Query Match 15.5%; Score 46.2; DB 24; Length 489;
Best Local Similarity 72.3%; Pred. No. 0.00038;
Matches 60; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 216 CATCTAGGAGCAAAATGAACCTCAATGCAGATCTAGCCAAACGTAAGATGAGACACA 275
DB 1 CATCTRAAGACATGATGAAGTCTGATTCGAATGCTAGTCATCATATAAATACCAAGACACA 60
QY 276 GAAATCGCTGAAACCTCTATTA 298
DB 61 GGTGCTCACTGACAAACTCTATCA 83

RESULT 10
AAT02405
ID AAT02405 standard; cDNA; 2646 BP.

XX AC AAT02405;
XX DT 30-JAN-1996 (first entry)
XX DE Rice mature pullulanase coding sequence.
XX KW Pullulanase; rice; brewing; low-calorie beer; high dextrose syrup;
XX KW yeast; Saccharomyces cerevisiae; ds.
XX OS Oryza sativa var. IR 36.

XX PN WO9509922-A1.
XX PD 13-APR-1995.
XX PF 04-OCT-1994; 94WO-US11242.
XX PR 05-OCT-1993; 93US-0132648.
XX PA (MILL-) MILLER BREWING CO.
XX PI Bower PA;
XX DR WPI; 1995-155263/20.
XX

Constructs contg. cloned rice pullulanase gene - for expression in yeast, useful in the brewing and beverage industries
Claim 8; Page 26-27; 43pp; English.

CC A rice flowering stage cDNA library was screened using a rice pullulanase genomic clone and a probe based on the pullulanase N-terminal sequence. cDNA clone 6-1 was obtd. This encoded a pullulanase lacking 13 N-terminal amino acids of the mature protein. The DNA sequence for these missing amino acids was determined from a PCR fragment and from a genomic clone. The resulting sequence for the mature pullulanase gene is given in CC AAT02399, and the complete coding sequence in AAT02405.

XX

SQ Sequence 2646 BP; 743 A; 509 C; 630 G; 764 T; 0 other;

Query Match 12.6%; Score 37.6; DB 16; Length 2646;
Best Local Similarity 55.3%; Pred. No. 0.29;
Matches 73; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 164 TGTTCTTCATATGATCTCTCATCGAGCGATTCAGATGATTAATGACAGAAATCATCTAGG 223
DB 285 TGTGCGCTTCTTATGATGCTCATGGAGGCGTCAAGATGTTACTGATTGCAACTACCTGG 344
QY 224 AGGACAAATGAACCTTCAATGCAGATTTAGCCAAACGTAAGATGAGACAGAAATCGC 283
DB 345 TGTATTGGATCACATGTTTGTCTTATACTGGACCACTTGGTGCAGTTTTTCACTGATAAGA 404
QY 284 TGAACACCTCTTA 295
DB 405 TGTGGACCTCTA 416

RESULT 11
AAT02400
ID AAT02400 standard; DNA; 2982 BP.

XX AC AAT02400;
XX DT 30-JAN-1996 (first entry)
XX DE Pullulanase expression construct.
XX KW Pullulanase; rice; brewing; low-calorie beer; high dextrose syrup;
XX KW yeast; Saccharomyces cerevisiae; ds.
XX OS Oryza sativa var. IR 36.
XX

XX FH Key Location/Qualifiers
XX FT misc_difference 2901
XX FT /*tag= a
XX FT /note= "base n at position 2901 is not identified
XX FT in the specification"
XX PN WO9509922-A1.
XX PD 13-APR-1995.
XX PF 04-OCT-1994; 94WO-US11242.
XX PR 05-OCT-1993; 93US-0132648.
XX PA (MILL-) MILLER BREWING CO.
XX PI Bower PA;
XX DR WPI; 1995-155263/20.
XX

Constructs contg. cloned rice pullulanase gene - for expression in yeast, useful in the brewing and beverage industries
Claim 3; Page 23-24; 43pp; English.
CC 5' 147 bp and 3' 701 bp end fragments of the rice pullulanase gene were combined with a 2.3 kb portion of pullulanase cDNA clone 6-1 (see: AAT02399) to obtain an expression construct (AAT02400) encoding a pullulanase lacking the first 2 amino acids of the native enzyme. The construct was used to express pullulanase in S. cerevisiae SEY2102 (ATCC 74281) for use in low-calorie beer or high dextrose syrup prodn.

Sequence 2982 BP; 875 A; 554 C; 690 G; 862 T; 1 other;

Query Match 12.6%; Score 37.6; DB 16; Length 2982;
Best Local Similarity 55.3%; Pred. No. 0.31;
Matches 73; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

OY 164 TGTTCCTCAATGATCCTCATCGAGGCGATTCAGATGATAATGACAGAAATCATCTAGG 223
 DB 279 TGTGCTCTTATGATGCTCATGGAGCGTCAAGATGTTACTGGATTGCAACTACCTGG 338
 OY 224 AGGACAATGACATTCATGAGATTCAGCCATGTAAGATGACAGAAATCGC 283
 DB 339 TGTATTGGATGACATGTTGCTTATGCTGACACCTTGGTGCAGTTCAGTGATAAAGA 398
 OY 284 TGAACACCTCTA 295
 DB 399 TGTGGACCTCTA 410
 RESULT 12
 AAT02399
 ID AAT02399 standard; cDNA; 2988 BP.
 XX
 AC AAT02399;
 XX
 DT 30-JAN-1996 (first entry)
 XX
 DE Rice mature pullulanase gene.
 XX
 KW Pullulanase; rice; brewing; low-calorie beer; high dextrose syrup;
 KW yeast; Saccharomyces cerevisiae; ds.
 XX
 OS Oryza sativa var. IR 36.
 XX
 FH Key Location/Qualifiers
 FT misc_difference 2907
 FT /*tag= a
 FT /note= "base n at position 2907 is not identified
 in the specification"
 XX
 PN - WO9509922-A1.
 XX
 PD 13-APR-1995.
 XX
 PF 04-OCT-1994; 94WO-US11242.
 XX
 PR 05-OCT-1993; 93US-01326648.
 XX
 PA (MILL-) MILLER BREWING CO.
 XX
 PI Bower PA;
 XX
 PT WPI; 1995-155263/20.
 XX
 PS Constructs contg. cloned rice pullulanase gene - for expression in
 PT yeast, useful in the brewing and beverage industries
 XX
 PS Disclosure; Page 21-22; 43pp; English.
 XX
 CC A rice flowering stage cDNA library was screened using a rice
 CC pullulanase genomic clone and a probe based on the pullulanase N-
 CC terminal sequence. cDNA clone 6-1 was obt'd. This encoded a
 CC pullulanase lacking 13 N-terminal amino acids of the mature
 CC protein. The DNA sequence for these missing amino acids was
 CC determined from a PCR fragment and from a genomic clone. The
 CC resulting sequence for the mature pullulanase gene is given in
 CC AAT02399.
 XX
 SQ Sequence 2988 BP; 876 A; 554 C; 695 G; 862 T; 1 other;
 Query Match 12.68; Score 37.6; DB 16; Length 2988;
 Best Local Similarity 55.38; Pred. No. 0.31; Mismatches 0; Gaps 0;
 Matches 73; Conservative 0; Indels 0; Gaps 0;
 OY 164 TGTTCCTCAATGATCCTCATCGAGGCGATTCAGATGATAATGACAGAAATCATCTAGG 223
 DB 285 TGTGCTCTTATGATGCTCATGGAGCGTCAAGATGTTACTGGATTGCAACTACCTGG 344
 OY 224 AGGACAATGACATTCATGAGATTCAGCCATGTAAGATGACAGAAATCGC 283

DB 345 TGTATTGGATGACATGTTGCTTATGCTGACCACTTGGTGCAGTTCAGTGATAAAGA 404
 OY 284 TGAACACCTCTA 295
 DB 405 TGTGGACCTCTA 416
 RESULT 13
 AAC48551
 ID AAC48551 standard; DNA; 4282 BP.
 XX
 AC AAC48551;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 57889.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126284.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.

PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149358.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139859.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			

Query Match 12.0%; Score 36; DB 21; Length 4282;

Best Local Similarity 49.0%; Pred. No. 1.1;

Matches 96; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

OY 85 ATGGATCGGCAANTAGGGTTTATTTATTTCTCAGCTACTAGTATTTTGTGGAGAAG 144

DB 809 ATGCATTGGACCAAGCAATAGAGGGTTAGCTGAGTTCTTATGATCGTCTTGAGGATG 868

PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.

PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 11-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-483232/52.
 XX Nucleic acids encoding 973 human testicular antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating testicular cancer
 XX Disclosure; SEQ ID NO 2279; 766pp; English.
 XX The present invention provides the protein and coding sequences of 973
 CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and
 CC gastrointestinal disorders, infections, and particularly cancer,
 CC especially testicular cancers. The present sequence is a DNA encoding a
 CC protein fragment of the invention.
 XX
 SQ Sequence 10859 BP; 3213 A; 2333 C; 2455 G; 2858 T; 0 other;
 Query Match 11.9%; Score 35.6; DB 23; Length 10859;
 Best Local Similarity 49.2%; Pred No. 2;
 Matches 125; Conservative 0; Mismatches 124; Indels 5; Gaps 1;
 Qy 25 CAGAAAATTCACCTTCGAGATGGCCTCTAAATCCTGGCTGAATTTTAACTTCTCT 84
 Db 7324 CAGTAATTTTCCCTTATCATGAATTTCTTGGGACTTTTACTTGGCTTTATAAACTCTCT 7383
 Qy 85 ATGGATCGGCAATAGGCTTTATTTATTTCTCAGCTACTAGTATTATTGTTGGGGAAG 144
 Db 7384 AGGCAATGCAATCGGCTTTCATGGCTTGGCTCTCTCTCTCTCTCTCTCTCTCTCT 7438
 Qy 145 AGGCTGACACCCAGCACTAATGTTCTTCAATATGATCTCATGCGAGGCATTCAGATGATA 204
 Db 7439 AGTTCCCTCGGTAGTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 7498
 Qy 205 ATGGACAGAAATCATCTTAGGAGGACAAATGAACCTTCAATGCAGATTTCTAGCCAAAG 264
 Db 7499 AAATTCATGTCATTCATTAGTTTAAAGGAATTTATCTCTATCTCTATCTCTGTTGCAAACT 7558
 Qy 265 ATGAGAACACAGAA 278
 Db 7559 GTGATATCAGATAA 7572

Search completed: June 27, 2003, 23:39:43
 Job time : 119.627 secs

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	37.6	12.6	2646	1	US-08-365-189-7	Sequence 7, Appli	
2	37.6	12.6	2982	1	US-08-365-189-2	Sequence 2, Appli	
3	37.6	12.6	2988	1	US-08-365-189-1	Sequence 1, Appli	
4	32.8	11.0	2042	2	US-08-911-434A-2	Sequence 2, Appli	
5	32	10.7	800	4	US-09-222-938A-68	Sequence 68, Appl	
6	31.8	10.6	1674	4	US-08-793-410-31	Sequence 31, Appl	
7	31.2	10.4	6216	4	US-09-415-522-5	Sequence 5, Appli	
8	31	10.4	1059	4	US-08-837-199A-23	Sequence 23, Appl	
9	31	10.4	1059	4	US-08-837-199A-47	Sequence 47, Appl	
10	30.8	10.3	1506	1	US-08-820-958-1	Sequence 1, Appli	
11	30	10.0	3541	4	US-09-180-439-5	Sequence 5, Appli	
12	30	10.0	3979	4	US-09-180-439-1	Sequence 1, Appli	
13	30	10.0	3979	4	US-09-180-439-2	Sequence 2, Appli	
14	30	10.0	4123	4	US-09-180-439-7	Sequence 7, Appli	
15	30	10.0	6471	4	US-09-353-585-1	Sequence 1, Appli	
16	29.8	10.0	12311	1	US-08-750-717-1	Sequence 1, Appli	
17	29.6	9.9	1371	4	US-09-134-001C-935	Sequence 935, App	
18	29.6	9.9	1893	4	US-09-134-001C-1778	Sequence 1778, Ap	
19	29.6	9.9	6816	4	US-09-404-650-1	Sequence 1, Appli	
20	29.6	9.9	6855	4	US-09-404-650-3	Sequence 3, Appli	
21	29.6	9.9	37948	4	US-09-251-645-11	Sequence 11, Appl	
22	29.4	9.8	1524	4	US-09-134-001C-1141	Sequence 1141, Ap	
23	29.4	9.8	4865	3	US-08-894-017-24	Sequence 24, Appl	
24	29.4	9.8	8920	2	US-08-446-855A-1	Sequence 1, Appli	
25	29.4	9.8	8920	4	US-09-150-741-1	Sequence 1, Appli	
26	29.4	9.8	111282	4	US-09-754-250-3	Sequence 3, Appli	
27	29.2	9.8	399	4	US-09-370-838-105	Sequence 105, App	


```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911.434A
; FILING DATE: 12-AUG-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Raymond C.
; REGISTRATION NUMBER: 21,066
; REFERENCE/DOCKET NUMBER: 2185-0199P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)205-8000
; TELEFAX: (703)205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2042 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORGANISM: Daucus carota L.
; INDIVIDUAL ISOLATE: Kuroda Gosun
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..2042
; US-09-911-434A-2

Query Match
Best Local Similarity 11.0%; Score 32.8; DB 2; Length 2042;
Matches 97; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

Qy 12 CCAGGCGGATGACAGAAAATTCACCTTCCGAGATGCCCTCTAAATCCTGCTGAATTTT 71
Db 197 CCACCGGTAATGGAAGATATGCAAAAAATTCAGCCACAGAGTCTTACTTAACCTCT 256
Qy 72 TTRACCTTCTCTATGATCGGCAATAGGGTTTATTTTATTTTCTCAGCTACTTAGTATT 131
Db 257 CACCTTTTGACACACTTTCTCATATTTTATTTTATTTTGTTCAAAAAATTTGAATA 316
Qy 132 TTGTGGGAGAAGAGGTCACCCAGACTAATGCTTCTCATATGATCTCTCATGCGAGG 191
Db 317 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 376
Qy 192 CATTGATGATGATGACAGAAAT 215
Db 377 AATAATAATAATAATAATAATAAT 400

RESULT 5
US-09-222-938A-68
; Sequence 68, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222.938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 68
; LENGTH: 800
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:

;
; NAME/KEY: CDS
; LOCATION: (114)...(743)
; US-09-222-938A-68

Query Match
Best Local Similarity 10.7%; Score 32; DB 4; Length 800;
Matches 86; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Qy 122 ACTTAGTATTTTGGGAGAAGAGGTGACACCCAGACTAATGCTTCTCATATAATGATCC 181
Db 233 AATTAGTAACCTTACACGACACACAAAGAGCCCAAAACAAAGTTGACCAAAATTCAGGA 292
Qy 182 TCATGCCAGGCATTTCAGATGATATGACAGAGATCATCTAGGAGGACAAAATGAACTTCAA 241
Db 293 GCAAGTATCAGCTATTCAAGCTGAGCAGTCTAACTTGAAGCTGAAAATGATAGATTACA 352
Qy 242 TGCAGATTCTAGCCAACTGTAAGATGAGACACAGAAATCGCTGAAAACCTCTATT 297
Db 353 AGCAGAATCTAAGAAACTCGAGGCTGAGATTACAGAACTTTCTAAAAACATTTGTT 408

RESULT 6
US-08-793-410-31
; Sequence 31, Application US/08793410
; Patent No. 5955650
; GENERAL INFORMATION:
; APPLICANT: HITZ, WILLIAM DEAN
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF CANOLA
; TITLE OF INVENTION: AND SOYBEAN PALMITYL-ACP THIO-
; TITLE OF INVENTION: ESTERASE GENES AND THEIR USE IN
; TITLE OF INVENTION: THE REGULATION OF FATTY ACID
; TITLE OF INVENTION: CONTENT OF THE OILS OF SOYBEAN
; TITLE OF INVENTION: AND CANOLA PLANTS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793.410
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10627
; FILING DATE: AUGUST 25, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CHRISTENBURY, LYNNE M.
; REGISTRATION NUMBER: 30,971
; REFERENCE/DOCKET NUMBER: CR-9567-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-5481
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1674 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; US-08-793-410-31

Query Match
Best Local Similarity 10.6%; Score 31.8; DB 2; Length 1674;
Matches 75; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
```


Sequence 1, Application US/08220958
Patent No. 5459072
GENERAL INFORMATION:
APPLICANT: McKay, Larry
TITLE OF INVENTION: FOOD-GRADE INTEGRATION VECTORS FOR
INDUSTRIAL BACTERIAL STRAINS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: MERCHANT & GOULD
STREET: 3100 NO. 5459072 West Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/220,958
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/841,642
FILING DATE: 25-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Kowalchuk, Alan W.
REGISTRATION NUMBER: 31,535
REFERENCE/DOCKET NUMBER: 600,229-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1506 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Lactococcus lactis subsp cremoris
STRAIN: SK11
IMMEDIATE SOURCE:
CLONE: ORI Site/OPEN Reading Frame of Replication
CLONE: Region of PSK11L
POSITION IN GENOME:
UNITS: bp
FEATURE:
NAME/KEY: CDS
LOCATION: 345..1496
US-08-220-958-1

Query Match 10.3%; Score 30.8; DB 1; Length 1506;
Best Local Similarity 50.7%; Pred. No. 1.6;
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 153 ACCGACAAATGTTCTTCATATGATCTCTGCGAGGATTTCAGATGATGAGAC 212
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1041 ACTCATTTTAATGTTGAGTATAAAAAATAAAAAAGGAGTTCAATCGATTCAATACAG 1100
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 213 AATCATCTAGGAGGACAAATGAACTCAATGCGAGATTCAGCCCAAGTAAAGATGAGAAC 272
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1101 TTTCATATAGTTAAGAACAAATTTGGAAGATGAGAACTATAAATGATGATGACAA 1160
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 273 ACAGAAATCGCTGAAACCTCTATTA 298
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1161 GCACAATTAACCTGAACAACAAATCA 1186

RESULT 11

US-09-180-439-5/c
Sequence 5, Application US/09180439
Patent No. 6225532
GENERAL INFORMATION:
APPLICANT: Dixon, Mark S
TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jones, Jonathan DG
STREET: 620 - 53
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/180,439
FILING DATE:
CLASSIFICATION: 3541
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/841,642
FILING DATE: 25-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Kowalchuk, Alan W.
REGISTRATION NUMBER: 31,535
REFERENCE/DOCKET NUMBER: 600,229-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1506 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Lycopodium obscurum
STRAIN: SK11
IMMEDIATE SOURCE:
CLONE: ORI Site/OPEN Reading Frame of Replication
CLONE: Region of PSK11L
POSITION IN GENOME:
UNITS: bp
FEATURE:
NAME/KEY: CDS
LOCATION: 345..1496
US-09-180-439-1

Query Match 10.0%; Score 30; DB 4; Length 3541;
Best Local Similarity 49.4%; Pred. No. 4.1;
Matches 78; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 67 ATTTTAACTTCTCTATGATCGGCAATAGGGTTTATTTATTTCTCAGCTACTTA 126
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 294 ATTTTCAATCTTCTTTGTCATGTTGAGAGGGTATTTATTTATGAGGACTACTTG 235
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 127 GTATTTTGTGGAGAGAGGGTGACACCCAGACTAATGTTCTTCATAATGATCCTCATG 186
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 234 ATATTTGATGCTTATGATGTGGAACCTTTTGAAATGAATATTCAAATAATGAGAGC 175
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 187 CGAGGCATTCAGATGATTAATGGACAGAAATCATCTAGGA 224
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 174 AAGAATATCAGAGATAATGTCAGAAATTAATTTGA 137
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-09-180-439-1/c
Sequence 1, Application US/09180439
Patent No. 6225532
GENERAL INFORMATION:
APPLICANT: Dixon, Mark S
TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jones, Jonathan DG
STREET: 620 - 53
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/180,439
FILING DATE:
CLASSIFICATION: 3541
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/841,642
FILING DATE: 25-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Kowalchuk, Alan W.
REGISTRATION NUMBER: 31,535
REFERENCE/DOCKET NUMBER: 600,229-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-9081
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1506 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Lycopodium obscurum
STRAIN: SK11
IMMEDIATE SOURCE:
CLONE: ORI Site/OPEN Reading Frame of Replication
CLONE: Region of PSK11L
POSITION IN GENOME:
UNITS: bp
FEATURE:
NAME/KEY: CDS
LOCATION: 345..1496
US-09-180-439-1

Query Match 10.0%; Score 30; DB 4; Length 3979;
Best Local Similarity 49.4%; Pred. No. 4.3;
Matches 78; Conservative 0; Mismatches 80; Indels 0; Gaps 0;


```

Search completed: June 28, 2003, 00:23:59
Job time : 27.4226 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic : nucleic search, using sw model

Run on: June 27, 2003, 23:29:41 ; Search time 77.4061 Seconds
(without alignments)
5734.045 Million cell updates/sec

Title: US-10-029-359A-1

Perfect score: 299

Sequence: 1 gcaagaggagccacgcccgcg.....tcgctgaaacacctctattag 299

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	299	100.0	299	9	US-10-029-359A-1
2	236.4	79.1	1794	9	US-10-235-056-2
3	231	77.3	1092	9	US-10-235-056-12
4	194	64.9	1469	9	US-10-235-056-6
5	182.6	64.4	1092	9	US-10-235-056-14
6	189.2	63.3	1440	9	US-10-235-056-4
7	187.8	62.8	1092	9	US-10-235-056-13
8	91	30.4	1560	9	US-10-029-359A-5
9	66.2	22.1	555	9	US-10-198-846-13278
10	46.2	15.5	489	9	US-10-029-359A-3
11	35.6	11.9	10859	9	US-09-764-891-7408
12	35.2	11.8	879	9	US-10-198-846-6422
13	33.8	11.3	618	9	US-10-015-219-96
14	33.8	11.3	618	10	US-09-777-564-96
15	33.8	11.3	18692	9	US-10-092-154-1682
16	33.8	11.3	18692	10	US-09-764-847-1682
17	32.8	11.0	1919	10	US-09-864-761-10504
18	32.6	10.9	1975	10	US-09-864-761-15628
19	32.4	10.8	3996	9	US-10-087-464-42

c 20	32.4	10.8	4095	9	US-09-938-842A-2642	Sequence 2642, Ap
c 21	32	10.7	408	10	US-09-925-300-351	Sequence 351, App
c 22	32	10.7	800	9	US-10-154-251-68	Sequence 68, Appl
c 23	32	10.7	800	9	US-10-154-251-69	Sequence 69, Appl
24	31.8	10.6	640681	10	US-09-790-988-1	Sequence 1, Appli
25	31.4	10.5	470	10	US-09-833-381-1312	Sequence 1312, Ap
26	31.4	10.5	1957	10	US-09-864-761-3787	Sequence 3787, Ap
c 27	31.2	10.4	490	9	US-10-060-036-2349	Sequence 2349, Ap
28	31.2	10.4	4023	9	US-09-291-417-17	Sequence 17, Appli
29	31.2	10.4	4141	9	US-10-303-683-7	Sequence 7, Appli
30	31.2	10.4	4380	9	US-10-303-683-8	Sequence 8, Appli
31	31	10.4	6944	9	US-10-172-086-111	Sequence 111, App
c 32	31	10.4	640681	10	US-09-790-988-1	Sequence 1, Appli
33	30.8	10.3	430	10	US-09-960-352-3616	Sequence 3616, Ap
c 34	30.8	10.3	12744	9	US-09-764-891-8228	Sequence 8228, Ap
c 35	30.6	10.2	233	10	US-09-983-965-2302	Sequence 2302, Ap
c 36	30.6	10.2	402850	9	US-09-844-653-5	Sequence 5, Appli
37	30.4	10.2	495	10	US-09-728-445-761	Sequence 761, App
c 38	30.4	10.2	13409	9	US-09-764-891-9601	Sequence 9601, Ap
c 39	30.4	10.2	13409	9	US-10-205-428-930	Sequence 930, App
c 40	30.2	10.1	425	9	US-09-918-995-35210	Sequence 35210, A
c 41	30.2	10.1	668	10	US-09-864-761-32135	Sequence 32135, A
c 42	30.2	10.1	1233	9	US-10-194-489-2	Sequence 2, Appli
43	30.2	10.1	1968	10	US-09-880-107-3711	Sequence 3711, Ap
44	30	10.0	313	9	US-10-015-219-550	Sequence 550, App
45	30	10.0	313	10	US-09-777-564-550	Sequence 550, App

ALIGNMENTS

RESULT 1

US-10-029-359A-1
; Sequence 1, Application US/10029359A
; Publication No. US20030100056A1
; GENERAL INFORMATION:
; APPLICANT: Attersand, Anneli
; TITLE OF INVENTION: Protein Cluster II
; FILE REFERENCE: 10806-156
; CURRENT APPLICATION NUMBER: US/10/029,359A
; CURRENT FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 299
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21)..(299)
; OTHER INFORMATION:
US-10-029-359A-1

Query Match	100.0%	Score 299;	DB 9;	Length 299;
Best Local Similarity	100.0%	Pred. No. 3.le-83;		
Matches 299;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GCAAGAGGAGCCACGCCGATGACAGAAAATTCACITTCGCGAGATGCCCTCTAAATCCT	60	
Db	1	GCAAGAGGAGCCACGCCGATGACAGAAAATTCACITTCGCGAGATGCCCTCTAAATCCT	60	
Qy	61	GGCTGAATTTTAAACCTTCTCTATGATCGGCAATAGGCTTTATTTATTTCTCAGC	120	
Db	61	GGCTGAATTTTAAACCTTCTCTATGATCGGCAATAGGCTTTATTTATTTCTCAGC	120	
Qy	121	TACTTAGTATTTCTTGGGAGAGGGTGACACCCAGACTAATGTTCTTCATATGATC	180	
Db	121	TACTTAGTATTTCTTGGGAGAGGGTGACACCCAGACTAATGTTCTTCATATGATC	180	
Qy	181	CTCATGGAGGCATTCAGATGATATGACAGATCATCTAGGAGGACAAATGAATTC	240	
Db	181	CTCATGGAGGCATTCAGATGATATGACAGATCATCTAGGAGGACAAATGAATTC	240	

QY 241 ATGCAGATTCTAGCCACGTAAGATGAGAACACAGAAATCGCTGAAACCTCTATTAG 299
|||||
Db 241 ATGCAGATTCTAGCCACGTAAGATGAGAACACAGAAATCGCTGAAACCTCTATTAG 299

RESULT 2

US-10-235-056-2
; Sequence 2, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1794
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-235-056-2

Query Match 79.1%; Score 236.4; DB 9; Length 1794;
Best Local Similarity 90.8%; Pred. No. 2.5e-63;
Matches 252; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
QY 22 TGACAGAAATTCACATTTCGAGATGGCCCTCTAAATCCTGGCTGAATTTTAAACCTTC 81
Db 40 TTACAGAAATACACTTCGGAATGGCCCTAAATCCTGGCTGAATTTTAAACCTTC 99
QY 82 TCTATGGATGGCAATAGGTTTATTTATTTCTCAGCTACTTAGTATTTTGTGGAG 141
Db 100 TCTGTGGATCAGCAATAGGATTTCTTTATGTCTCAGCTATTAGTATTTTGTGGAG 159
QY 142 AACAGGTGACACCCAGCAATTAATCTTCTCATATATCTCCTCATGCGGCAATTCAGATG 201
Db 160 AAAAGTTGACACCCAGCAATTAATCTTCTCATATATCTCCTCATGCGGCAATTCAGATG 219
QY 202 ATAATGACAGAAATCTATAGGAGCAAAATGAATCTCAATGAGATTTCTAGCCACGTA 261
Db 220 ATAATGACAGATCTATAGAGGACAAATGAATCTCAATGAGATTTCTAGCCACATA 279
QY 262 AAGATGAGACACAGAAATCGCTGAAACCTCTATTAG 299
Db 280 AAGATGAGACACAGCAATTCGCTGAAACCTCTATTAG 317

RESULT 3

US-10-235-056-12
; Sequence 12, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1092
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-235-056-12

Query Match 77.3%; Score 231; DB 9; Length 1092;
Best Local Similarity 94.1%; Pred. No. 9.4e-62;
Matches 240; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 45 ATGCGCTCTAAATCCTGGCTGAATTTTAAACCTTCTCTATGGATCGGCAATAGGTTT 104
Db 1 ATGCGCTCTAAATCCTGGCTGAATTTTAAACCTTCTCTCTGTGGATCAGCAATAGGATT 60
QY 105 ATTTTATTTTCTCAGCTACTTAGTATTTTGTGGGAGAAGAGGTGACACCCAGCAAT 164
Db 61 CTTTATGTCTCAGCTACTTAGTATTTTGTGGGAGAAGAGGTGACACCCAGCAAT 120
QY 165 GTTCTTCATATGATCCTCATGCGGCAATTCAGATCATATGACAGAAATCATCTAGGA 224
Db 121 GTTCTTCATATGATCCTCATGCGGCAATTCAGATCATATGACAGAAATCATCTAGGA 180
QY 225 GGACAAATGAATTCATGAGATTTCTAGCCAACTGAAGATGAGAACAGAAATCGCT 284
Db 181 GGACAAATGAATTCATGAGATTTCTAGCCAACTGAAGATGAGAACAGAAATCGCT 240
QY 285 GAAACCTCTATTAG 299
Db 241 GAAACCTCTATTAG 255

RESULT 4

US-10-235-056-6
; Sequence 6, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1469
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-235-056-6

Query Match 64.9%; Score 194; DB 9; Length 1469;
Best Local Similarity 84.5%; Pred. No. 3.7e-50;
Matches 218; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 42 GAGATGGCTCTAAATCCTGGCTGAATTTTAAACCTTCTCTATGGATCGGCAATAGG 101
Db 177 GAAATGGCTCTAAATCCTGGCTGAATTTTAAACCTTCTCTCTGTGGATCAGCAATAGG 236
QY 102 TTTTATTTTATTTCTCAGCTACTTAGTATTTTGTGGGAGAAGAGGTGACACCCAGCT 161
Db 237 TTTTATTTTATTTCTCAGCTACTTAGTATTTTGTGGGAGAAGAGGTGACATTCAGCCT 296
QY 162 AATGTTCTTCATATGATCCTCATGCGGCAATTCAGATCATATGAGAACAGAAATCATCTA 221
Db 297 AACATGCTTCACATGACCTCATGCAAGGCAATTCAGATCATATGAGAACAGAAATCATCTC 356
QY 222 GGAGGCAATGAACTTCAATGAGATTTCTAGCCAACTGAAGATGAGAACAGAAATC 281
Db 357 AAAGGACAGATGAATTCATGAGATTTCTAGCCAACTGAAGATGAGAACAGAAATC 416
QY 282 GCTGAAACCTCTATTAG 299
Db 417 GCTGAGAACCTCTATTAG 434


```

US-10-092-154-1682
; Sequence 1682, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1682
; LENGTH: 18692
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1682

Query Match          11.3%; Score 33.8; DB 9; Length 18692;
Best Local Similarity 56.9%; Pred. No. 11;
Matches 62; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 139 GAGAAGAGGGTGACACCCAGACTTAATGTTCTTCATATGATCTCATGCGAGGCATTTCAG 198
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3999 GTCAATTTCATCTGTCGACCCCTTAGCCCTTAAGAAAGATCTTCAGCCCGCAGTCAG 4058

QY 199 ATGATTAATGGACAGATCATCTATAGGAGGACAAATGAACTTCAATGCAGA 247
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4059 ATAATGATGTTTCGGAATAGCATATCAGGAATGCAGAAATTCATGGGAGA 4107
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: June 28, 2003, 00:27:34
Job time : 78.4061 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 23:34:12 ; Search time 753.192 Seconds
(without alignments)
6429.241 Million cell updates/sec

Title: US-10-029-359a-1

Perfect score: 299

Sequence: 1 gcaagaggagcagcgccg.....tcgtgaaacacctattag 299

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vit:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	198.2	66.3	462	9 AI236007	AI236007 EST232569
2	189.2	63.3	471	9 AI059600	AI059600 UI-R-C1-1
3	182	60.9	594	12 BF471960	BF471960 UI-M-BH3-
4	153.6	51.4	316	12 BE993641	BE993641 UI-M-BZ1-
5	130	43.5	642	9 AJ447748	AJ447748 AJ447748
6	93.4	31.2	614	9 AJ449471	AJ449471 AJ449471

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

7	83.4	27.9	570	9	AL698324	AL698324 DKF2P686P
8	67.6	22.6	245	14	BQ337716	BQ337716 CM3-NN002
9	67.2	22.5	618	17	AZ456919	AZ456919 IM0260104
10	60.6	20.3	481	12	BF443807	BF443807 261354 MA
11	60	20.1	141	14	T10488	T10488 hbc402 Huma
12	40.4	13.5	701	17	AZ581501	AZ581501 IM0370M15
13	39.6	13.2	632	12	BG447948	BG447948 NF104B09E
14	38	12.7	695	13	BI667399	BI667399 603292180
15	38	12.7	696	13	BI102329	BI102329 602885661
16	37.2	12.4	541	17	CNS04KUS	AL295309 Tetraodon
17	37	12.4	445	12	BG293789	BG293789 602390685
18	37	12.4	712	17	AZ176339	AZ176339 SP_0141-A
19	36.8	12.3	451	17	AQ337232	AQ337232 HS_5009-A
20	36.8	12.3	669	17	CNS010M7	AL099145 Drosophila
21	36.6	12.2	428	14	BQ452715	BQ452715 sao90e05
22	36.4	12.2	369	13	BI506540	BI506540 BB170029A
23	36.4	12.2	709	13	BJ349463	BJ349463 BJ375529
24	36.4	12.2	896	13	BJ375529	BJ375529 BJ375529
25	36.2	12.1	480	10	AW662066	AW662066 h118a03..x
26	36.2	12.1	530	9	AI911553	AI911553 ty73601.x
27	36	12.0	524	17	B00008	B00008 CSRL-100b5-
28	36	12.0	678	9	AU237109	AU237109 AU039635
29	35.8	12.0	537	9	AU039635	AU039635 AU039635
30	35.8	12.0	540	9	AU034415	AU034415 AU034415
31	35.8	12.0	769	13	BJ407722	BJ407722 BJ407722
32	35.8	12.0	778	12	BG603451	BG603451 EST502541
33	35.8	12.0	790	13	BJ403177	BJ403177 BJ403177
34	35.6	11.9	522	14	C25513	C25513 C25513 Dict
35	35.6	11.9	614	9	AU060072	AU060072 AU060072
36	35.6	11.9	784	17	BH493958	BH493958 BOHRK43TF
37	35.6	11.9	836	17	AZ185234	AZ185234 SP_1004..B
38	35.6	11.9	852	14	BQ717911	BQ717911 AGENCOURT
39	35.6	11.9	1101	17	CNS00EHW	AL069213 Drosophila
40	35.4	11.8	686	13	BJ361849	BJ361849 BJ361849
41	35.2	11.8	368	12	BG526841	BG526841 40-45 Ste
42	35.2	11.8	578	14	BP007292	BP007292 BP007292
43	35.2	11.8	717	10	AV896012	AV896012 AV896012
44	35	11.7	904	17	BH161018	BH161018 ENTRV85TF
45	34.8	11.6	283	13	BMI57689	BMI57689 NXLV_022..

ALIGNMENTS

RESULT 1
AI236007
LOCUS EST232569 Normalized rat ovary, Bonto Soares Rattus sp. cDNA clone
DEFINITION ROVDAL5 3' end, mRNA sequence.
ACCESSION AI236007
VERSION AI236007.1 GI:3829513
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE 1 (bases 1 to 462)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1..462
/organism="Rattus sp."

```

/db_xref="ATCC (inhost):2041425"
/db_xref="taxon:10118"
/clone="ROVDA15"
/clone_lib="Normalized rat ovary; Vector: p1773Pac; Site_1: EcoRI;
Site_2: NotI"
BASE COUNT 121 a 113 c 117 g 111 t
ORIGIN

Query Match 66.3%; Score 198.2; DB 9; Length 462;
Best Local Similarity 78.9%; Pred. No. 1.9e-43;
Matches 236; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1 GCAAGAGGAGCCAGCGCGATGACAGAAATTCACATTTCCGAGATGGCCTCTAAATCCT 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
98 GCGAGAGAGCGCGCGGTGATGTTCAGAAATACATTTCCAGGAATGCCCTCTAAATCTT 157

QY 61 GCGTGAATTTTAACTTCTCTATGATCGGCAATAGGGTTTATTTATTTCTTCAGC 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
158 GCGTGAATTTTAACTTCTCTGAGGATCAGCAATAGGATTTTCTTATGTTCTCAAC 217

QY 121 TACTTAGTATTTTGGGAGAGGGTGACACCCAGACTAATCTTCTTCATATGATC 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
218 TCTTGAATTTCTGTGCAAGACAGCGTGACGTTACGCTTACATGCTTCAATGATC 277

QY 181 CTCATCGGAGGATTCAGATGATATGACAGAAATCATCTTAGGAGGACAAATGAACCTTCA 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
278 CTCATCGGAGGATTCAGACGACAGTGGACATAATCACCTCAAGGACAGATGGACTTCA 337

QY 241 ATGCAGATTCAGCCACGTAAGATGAGACACAGAAATCCTGAAACCTCTATTAG 299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
338 ATGCAGATTCATCAACATACAGATGAGACACAGACGTTGCTGAGAACCTCTATCAG 396

RESULT 2
AIO59600
LOCUS
DEFINITION
UI-R-C1-lj-g-05-0-UI.s1 UI-R-C1 Rattus norvegicus cDNA clone
UI-R-C1-lj-g-05-0-UI 3', mRNA sequence.
AIO59600
ACCESSION
VERSION
AIO59600.1 GI:3333377
KEYWORDS
EST.
SOURCE
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 471)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
9704477
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.
Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
through Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LNL (info@image.lnl.gov). IMAGE
ID=1784976
Seq primer: M13 Forward
POLYA-No.
Location/Qualifiers
1..471
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"

FEATURES
source
LOCUS
BF471960
DEFINITION
UI-M-BH3-awu-c-04-0-UI.r1 NIH_BMAP_M.S4 Mus musculus cDNA clone
UI-M-BH3-awu-c-04-0-UI 5', mRNA sequence.
ACCESSION
VERSION
BF471960.1 GI:11541143
KEYWORDS
EST.
SOURCE
house mouse.
Mus musculus.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 594)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
9704477
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.
Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
through Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LNL (info@image.lnl.gov). IMAGE
ID=1784976
Seq primer: M13 Forward
POLYA-No.
Location/Qualifiers
1..471
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"

/clone="UI-R-C1-lj-g-05-0-UI"
/clone_lib="UI-R-C1"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: p1773D-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C1
library is a subtracted library derived from the UI-R-C0
library, which is a subtracted library derived from the
UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
consisted of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung,
brain, liver, kidney, heart, spleen, ovary, and muscle.
The UI-R-E1 library consisted of a mixture of
individually tagged normalized libraries constructed from
8, 12 and 18-day embryo. The tag is a string of 3-5
nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-C1) was constructed as follows: PCR
amplified cDNA inserts from UI-R-C0 clones from which 3'
ESTs had been derived was used as a driver in a
hybridization with the UI-R-C0 library in the form of
single-stranded circles. The remaining single-stranded
circles (subtracted library) was purified by
hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-C1
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)."
BASE COUNT 122 a 117 c 101 g 131 t
ORIGIN

Query Match 63.3%; Score 189.2; DB 9; Length 471;
Best Local Similarity 83.3%; Pred. No. 5.3e-41;
Matches 215; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 42 CAGATGGGCTCTAAATCCTGCTGAATTTTAACTTCTCTATGGATCGCAATAGGG 101
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
169 GAAATGGGCTCTAAATCCTGCTGAATTTTAACTTCTCTATGGATCGCAATAGGA 228

QY 102 TTTATTTATTTTCTCAGCTACTAGTATTTTGGGAGAGAGGGTGACACCCAGACT 161
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
229 TTTTCTTATGTTCTCAACTCTTGAATATTTTGTGCAAGACAGCGCTTCCAGCT 288

QY 162 AATGTTCTTCAATGATCCTCATCGAGGATTCAGATGATTAATGACAGATCATCTA 221
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
289 AACATGCTTCATATGATCCTCATCGAGATTCAGACAGATGGACATATCACCTC 348

QY 222 GGAGGACAAATGAATTCATGAGTCTAGCCCACTAAAGATGAGAACACAGAAATC 281
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
349 AAAGCACAGATGGACTTCAATGCAGATCTAGCCCACTAAAGATGAGAACACAGACGTT 408

QY 282 CCGTGAACACCTCTATTAG 299
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
409 GCTGAGAACCTCTATCAG 426

RESULT 3
BF471960/c
LOCUS
DEFINITION
UI-M-BH3-awu-c-04-0-UI.r1 NIH_BMAP_M.S4 Mus musculus cDNA clone
UI-M-BH3-awu-c-04-0-UI 5', mRNA sequence.
ACCESSION
VERSION
BF471960.1 GI:11541143
KEYWORDS
EST.
SOURCE
house mouse.
Mus musculus.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 594)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
9704477
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.
Fatima Bonaldo, Ph.D. Clone distribution: clones will be available
through Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LNL (info@image.lnl.gov). IMAGE
ID=1784976
Seq primer: M13 Forward
POLYA-No.
Location/Qualifiers
1..471
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"

```



```

QY 192 CATTGAGATGAATGGACAGAAATCATCTAGGAGGACAAATGAACCTTCAATGCAGATTCT 251
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 129 CATTGAGATGAATGGACAGAGTCACCTCAAAAGGACAGATGAACCTTCAATGCAGATTCC 188
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 252 ACCCAACGTAAAGATGAGACACAGAAATCGCTGAAACCTCTATTAG `299
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 189 ACCCAACATAAAGATGAGAACATAGACGTTGCTGAGAACCTCTATCAG 236
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
AJ447748
LOCUS AJ447748 642 bp mRNA linear EST 19-APR-2002
DEFINITION Gallus gallus bursal CDNA clone 1717r1, mRNA sequence.
ACCESSION AJ447748
VERSION AJ447748.1 GI:20214969
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 642)
Buerstedde,J.M.
Gallus gallus bursal lymphocyte EST
Unpublished (2002)
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES
source
Location/Qualifiers
1..642
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="1717r1"
/clone_lib="rikeni"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/note="CB inbred strain"
BASE COUNT 174 a 135 c 172 g 158 t 3 others
ORIGIN
Query Match 43.5%; Score 130; DB 9; Length 642;
Best Local Similarity 69.7%; Pred. No. 6.5e-25;
Matches 175; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 49 CCTCTAAATCCTCGCTGAATTTTTTAACTTCCTCTATGGATCGGCAATAGGGTTTATT 108
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 145 CTTCTAAGTCTTGTAGAACCTCTCGACCTTCAGTTTGGATCTGCAATAGGATTTTTC 204
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 109 TATTTTCTCAGCTACTTAGTATTTTGGGAGAGAGGGTGACACCCAGACTAATGTTTC 168
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 205 TATGATATCTCTGTTTGTAGATGATACAGAGAACAGGTTGAGATCAACCTCTATATTC 264
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 169 TTCAATGATCTCTCATCGAGGCAATTCAGATGATATGAGAGAAATCATCTAGAGGAGC 228
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 265 TTCAATGATCTCTCATCGGCAATTCAGAGATGATGATAAACAATCAGCTGCAAGGAC 324
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 229 AAATGAATTTCAATGCAGATTTCTAGCCCAACGTAAGATGAGAACACAAATCGCTGAAA 288
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 325 AAATGAATTTCAATGCAGATTTCTAGCCCAACGTAAGATGAGAACACAAATCGCTGAAA 384
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 289 ACCTCTATTAG 299
||||| ||||| |||||
Db 385 GACTCTATCAG 395
||||| ||||| |||||

RESULT 6
AJ449471
LOCUS AJ449471 614 bp mRNA linear EST 19-APR-2002
DEFINITION Gallus gallus bursal CDNA clone 22c13r1, mRNA sequence.
ACCESSION AJ449471

```

```

VERSION AJ449471.1 GI:20216692
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 614)
Buerstedde,J.M.
Gallus gallus bursal lymphocyte EST
Unpublished (2002)
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES
source
Location/Qualifiers
1..614
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="22c13r1"
/clone_lib="rikeni"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/note="CB inbred strain"
BASE COUNT 185 a 125 c 164 g 140 t
ORIGIN
Query Match 31.2%; Score 93.4; DB 9; Length 614;
Best Local Similarity 69.4%; Pred. No. 5.7e-15;
Matches 127; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 117 CAGCTACTTAGTATTTTGTGGGAGAGGGGTGACACCCAGACTAATGTTCTTCTAAT 176
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98 CGCGCGCTGAGCATGATACCTAGAGAACAGGTTGAGATCCCAACCTCATATTCTTCAAT 157
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 177 GATCCTCATGCGAGGATTCAGATGATATGACAGAAATCATCTAGGAGGACAAATGAAC 236
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 158 GATCCCCATGGTCAACACTCAGACATGATACATCAGCTGCAAGGACAAATGAAT 217
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 237 TTCAATGAGATTTCTAGCCCAACGTTAAAGATGAGACACAGAAATCGCTGAAACCTCTAT 296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 218 TTCAATGCGGACTCTGGACAGCATAGATGAGATGAGAAACATTCGACATGCATCTAT 277
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 297 TAG 299
|||
Db 278 GAG 280
|||

RESULT 7
AL698324
LOCUS AL698324 570 bp mRNA linear EST 21-MAR-2002
DEFINITION DKFZp686P14107_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DKFZp686P14107 5', mRNA sequence.
ACCESSION AL698324
VERSION AL698324.1 GI:19618864
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 570)
Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann,S.).
EST (Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann
,S.).
Unpublished (1999)
Contact: Poustka A.J.
Department Leirach
Max-Planck-Institute for Molecular Genetics
Innestrasse 73, 14195 Berlin, Germany
Tel: +49-30-84131623
Fax: +49-30-84131128
Email: poustka@mpimg-berlin-dahlem.mpg.de

```


/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCIM0260104"
/clone_lib="Mouse 10kb plasmid UUC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42uv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 193 a 137 c 126 g 161 t 1 others
ORIGIN

Query Match 22.5%; Score 67.2; DB 17; Length 618;
Best Local Similarity 57.3%; Pred. No. 7.5e-08;
Matches 164; Conservative 0; Mismatches 113; Indels 9; Gaps 2;

Qy 14 ACGCCCATGACAGAAATTCACCTTCGAGATGGCTCTAAATCCTGGCTGATTTTT 73
Db 181 ACTCCCTTGGTGGCAGAAATATCTCCCAAGCTTCCCATATTTCTGGTGCATTTCT 240

Qy 74 AACCTTCTCTATGATCGGCAATAGGTTTATTTATTTCTCAGCTACTAGTATTTT 133
Db 241 AGCTCCTCTCGAGATCAGTCATGATTTCTTTCTGTCAGCATTTACTGTGT 300

Qy 134 GTTGGAGAGAGGGTGACACCCAGACATAATGTTCTTCAATGATCCTCATCGAGGCA 193
Db 301 ACTGGAAGAAAGAAATGAC-----ATTAACGCTATGGCTCCTAATCCACTGTGATTC 354

Qy 194 TTCAGATGAATGACAGACATATCTAGGAGGACAATGAACCTCAATGCAGATTCTAG 253
Db 355 TTCAATCATACAGAGAAATGTCAAAAGTATTGGTATGGCAAAATAAATACT-- 412

Qy 254 CCAACGTAAGATGACAGAACACAGAAATCGCTGAAACCTCTATTAG 299
Db 413 -GCACATAGTATGATGATAGAGAAATGGCTGTAAACTCTACCAG 457

RESULT 10
BF443807
LOCUS 481 bp mRNA linear EST 01-DEC-2000
DEFINITION 261354 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF443807
VERSION BF443807.1 GI:11503836
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 481)
Stoner, R.T., Heaton, M.P., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Bennett, G.A., Laegreid, W.W.
and Keele, J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCACGACG
Plate: 94 row: J column: 20
Seq primer: ATTTAGGTGACACTATAG.

FEATURES
source
1..481
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from testis, ovary, "
endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 167 a 81 c 112 g 121 t
ORIGIN

Query Match 20.3%; Score 60.6; DB 12; Length 481;
Best Local Similarity 88.0%; Pred. No. 4.6e-06;
Matches 66; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 225 GGCAAAATGAATCAATGATTCAGATTCAGCAAGCTAAAGATGAGAACACAGAAATCGCT 284
Db 1 GGACAGATAAACTTGGTGCAGATGCTAGCCACATAAAGATGAGAACACAGATCGCT 60

Qy 285 GAAACCTCTATTAG 299
Db 61 GAAACCTCTATCAG 75

RESULT 11
T10488
LOCUS 141 bp mRNA linear EST 29-NOV-1993
DEFINITION hbc402 Human pancreatic islet Homo sapiens cDNA clone hbc402 3' end,
mRNA sequence.
ACCESSION T10488
VERSION T10488.1 GI:390642
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 141)
Takeda, J., Yano, H., Eng, S., Zeng, Y. and Bell, G.I.
A molecular inventory of human pancreatic islets: sequence analysis
of 1000 cDNA clones
Hum. Mol. Genet. 2, 1793-1798 (1993)
94108427
Contact: Bell GI or Takeda J
HHMI
Univ. of Chicago
5841 S. Maryland Ave., MC1028, Chicago IL 60637
Tel: 3127029116
Fax: 3127020271
Email: g-bell@uchicago.edu
Seq primer: T7 primer.
Location/Qualifiers
1..141
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hbc402"

FEATURES
source
1..141
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hbc402"
/note="vector: Lambda ZAPII; Site_1: Eco RI; Site_2: Xho
I; mRNA was prepared from normal adult human islets. cDNA

was directionally synthesized from the Xho I in the vector to the EcoRI site. cDNA was size fractionated to remove sequences <1000 bp in size."

BASE COUNT 57 a 30 c 23 g 31 t
ORIGIN

Query Match 20.1%; Score 60; DB 14; Length 141;
Best Local Similarity 92.4%; Pred. No. 6.3e-06;
Matches 85; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY 208 GACAGATCATCTAGGAGGACAAATCAACTTCAATGCAGATTTCAGCAACGTAACATG 267
|||||
Db 1 GACAGATCATCTAGGAGGACAAATCAACTTCAATGCAGATTTCAGCAACGTAACATG 59
|||||

QY 268 AGACACAGAAATCGCTGAAACCTCTATTAG 299
|||||
Db 60 AGACACAGACAT-TCTGAAACCTCTATCAG 90
|||||

RESULT 12
AZ581501/c
LOCUS
DEFINITION 701 bp DNA linear GSS 13-DEC-2000
IM0370M15F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0370M15 F, DNA sequence.

ACCESSION AZ581501
VERSION
KEYWORDS
SOURCE GSS.
ORGANISM house mouse.

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 701)

TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL
COMMENT Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177

Insert Length: 10000 Std Error: 0.00
Plate: 0370 Row: M Column: 15
Seq primer: CGTTGTAACGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 701.
Location/Qualifiers
1. .701
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0370M15"
/clone_lib="Mouse, 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: pW042nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative

FEATURES
source

of pW042 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 240 a 171 c 77 g 213 t
ORIGIN

Query Match 13.5%; Score 40.4; DB 17; Length 701;
Best Local Similarity 51.7%; Pred. No. 1.5;
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 25 CAGAAAAATTCACCTTCGAGATGCGCTCTAAATCTGGCTGAATTTTAACTTCCTCT 84
|||||
Db 477 CAGGTGTTGGTTTCAAAATAATTAATCTCTTTCAAGGCAAAAGTATTTTCGATTGTG 418
|||||

QY 85 ATGGATCGCAATAGGTTTATTTTCTCAGCTACTTACTATTCTTGGGAGAAG 144
|||||

Db 417 ACTGAACACCAATATGAGGTGGTTGTATACAGACACTGTCAGTTCTTAGATGAAG 358
|||||

QY 145 AGGTCGACACCCAGACTAATGTTCTTCAATATGATCTCTATGCGAGGCAATTCAGATGA 202
|||||

Db 357 TGGTGGCACTCAGCAGAAATGTCAGACAGAGATGAGTGAGGAATTTATTITA 300
|||||

RESULT 13
BG447948

LOCUS
DEFINITION 632 bp mRNA linear EST 16-MAR-2001
NF104B09ECIF1075 Elicited cell culture Medicago truncatula CDNA
clone NF104B09EC 5', mRNA sequence.

ACCESSION BG447948
VERSION
KEYWORDS BG447948.1 GI:13366729
SOURCE EST.

ORGANISM barrel medic.
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

REFERENCE 1 (bases 1 to 632)
AUTHORS Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
Center for Medicago Genomics Research
Unpublished (2000)
Contact: Dixon RA
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 632 Std Error: 0.00
Plate: 104 row: B column: 09

Seq primer: TCACACAGGAACAGCTATGAC.
Location/Qualifiers
1. .632
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF104B09EC"
/clone_lib="Elicited cell culture"
/tissue_type="Cell cultures derived from root tissues"
/dev_stage="Cell suspensions were subcultured every 14
days. Cells were induced six days after subculture"
/note="Vector: Lambda Zap; Cells were induced with yeast
cell wall extracts equivalent to 50ug/ml glucose in the
final concentration. Samples were taken at 0.5, 1, 12 and
24 hours after induction. Equal amounts of RNA from each
time point were pooled and used for mRNA isolation."

FEATURES
source

BASE COUNT 179 a 108 c 130 g 212 t
ORIGIN

```

Query Match      13.2%; Score 39.6; DB 12; Length 632;
Best Local Similarity 55.6%; Pred. No. 2.4;
Matches 75; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 93 GCATAGGGTTATTTATTTCTCAGCTACTAGTATTGTTGGGAGAGGGGTGAC 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 GCACCAATATATGTCATGATTTCACACCCCATATATGACATGAGGAAAAGGCC 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 153 ACCGAGCAATCTCTTCATATGATCCTCATGCGGAGGCATTCAGATGATATGACAG 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 TCACAATTTTGGCTTTAGGAGGAGCCTCACTCAAGTTTTCAGAAGCTGANGTACAT 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 213 AATCATCTAGGAGGA 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 271 AACCAACCAGAAAGA 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
BI667399
LOCUS      BI667399
DEFINITION 603292180F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5311723 5',
            mRNA sequence.
ACCESSION  BI667399
VERSION     BI667399.1 GI:15581632
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 695)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-femail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11789 row: o column: 20
            High quality sequence stop: 691.
            Location/Qualifiers
                location=1..695
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5311723"
                /clone_lib="NIH_MGC_96"
                /tissue_type="hypothalamus"
                /lab_host="DH10B"
                /notes="Organ: brain; Vector: pBluescriptR (modified
                pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
                ); Oligo-dT primed using primer 5'-TTTTTTTATTTTATTTN-3',
                size-selected for average insert size 2.3 kb and
                normalized to Rof 5. This is a primary library enriched
                for full-length clones and constructed using the
                Cap-trapper method (Carninci, in preparation). Library
                constructed by M. Brownstein (NIMH/NHGRI, National
                Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT      217 a 119 c 159 g 200 t
ORIGIN

Query Match      12.7%; Score 38; DB 13; Length 695;
Best Local Similarity 51.8%; Pred. No. 6.5;
Matches 86; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 63 CTCGAATTTTACCTCTCTCATGATCGGCAATAGGGTTTATTTATTTCTCAGCTA 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 CTAGTTGTTTAACTCTCTCTTTTACATTTGGCTTTTGTTCCTAAATGTTCTCCAAGCTA 573
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 123 CTTAGTATTTTGTGGAGAAGAGGTGACACCCAGCAATGCTTCTTCATAATGATCCT 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 574 TTGTATGTTGGATTGCAGAAGAATTGTGCACGATGAATACTTTTTTTTAAATGTCATTA 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 183 CATGCCAGGCATTCAGATGATGATAATGACAGAAATCATCTAGGAGGAC 228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 634 TTACCAATATTGAGTGAAGCTAATTGTCAACTTTTATTAGGATTAC 679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
BI102329
LOCUS      BI102329
DEFINITION 60285661F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5040878
            5', mRNA sequence.
ACCESSION  BI102329
VERSION     BI102329.1 GI:14553222
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-femail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1112 row: j column: 15
            High quality sequence stop: 696.
            Location/Qualifiers
                location=1..696
                /organism="Mus musculus"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="IMAGE:5040878"
                /clone_lib="NCI_CGAP_Kid14"
                /lab_host="DH10B (T1 phage-resistant)"
                /note="Organ: kidney; Vector: pCMV-SPORT6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.75 kb. Constructed by Life
                Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT      245 a 136 c 147 g 168 t
ORIGIN

Query Match      12.7%; Score 38; DB 13; Length 696;
Best Local Similarity 82.4%; Pred. No. 6.5;
Matches 56; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

QY 234 AACTTCAATCCAGATTCTACCCCAACCTAAGATGAGAACAC--AGAAATCGCTGAAACC 291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 AACTTCAATGCGAGATTCCAGCCCAACATAAAGATGAGAAGCAGTAGACGTTGCTGAGAACC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 292 TCTATTAG 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 TCTATCAG 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: June 28, 2003, 01:00:53
Job time : 757.192 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 22:30:21 ; Search time 1591.73 Seconds
(without alignments)
8940.745 Million cell updates/sec

Title: US-10-029-359A-3
Perfect score: 489
Sequence: 1 catctaaaagactgatgaa.....gatatgtactgagcaagaa 489

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	478	97.8	153472	9	AC011179 Homo sapi
C 2	478	97.8	165720	9	AC025524 Homo sapi
C 3	307	62.8	930	9	AJ132443 Homo sapi
C 4	307	62.8	1092	6	AX175170 Sequence
C 5	307	62.8	1092	9	AJ243256 Homo sapi
C 6	307	62.8	1092	9	AJ278960 Homo sapi
C 7	307	62.8	1794	6	AX175160 Sequence
C 8	307	62.8	1794	9	AF155582 Homo sapi
C 9	300.6	61.5	1092	6	AX175171 Sequence
C 10	300.6	61.5	1405	10	AF157963 Rattus no
C 11	300.6	61.5	1440	6	AX175162 Sequence
C 12	292.6	59.8	1092	6	AX175172 Sequence
C 13	292.6	59.8	1438	10	AF157962 Sequence
C 14	292.6	59.8	1469	6	AX175164 Sequence
C 15	281	57.5	1708	10	BC025899 Mus muscu
C 16	279.8	57.2	172574	2	AC011121 Homo sapi
C 17	279.8	57.2	182823	9	AC016638 Homo sapi
C 18	279.8	57.2	217807	2	AC079469 Homo sapi
C 19	278.2	56.9	169898	2	AC016585 Homo sapi
C 20	278	56.9	115954	9	AC005532 Homo sapi
C 21	275	56.2	146790	2	AC118920 Rattus no
C 22	275	56.2	194038	2	AC095494 Rattus no
C 23	270.4	55.3	1499	9	BC003174 Homo sapi
C 24	244.8	50.1	123070	9	AC020656 Homo sapi
C 25	244.8	50.1	170227	2	AC019161 Homo sapi
C 26	244.8	50.1	307755	2	AC127894 Homo sapi
C 27	197.8	40.4	171529	2	AC023021 Homo sapi
C 28	181	37.0	183416	10	AL671269 Mouse DNA
C 29	181	37.0	229117	2	AL807383 Mus muscu
C 30	142	29.0	156337	2	AC009938 Homo sapi
C 31	116.2	23.8	167996	9	AC021753 Homo sapi
C 32	116.2	23.8	191655	9	AC020661 Homo sapi
C 33	115.6	21.5	1170	6	AX175173 Sequence
C 34	105	21.5	1172	3	AF269063 Caenorhab
C 35	105	21.5	1172	6	AX175166 Sequence
C 36	105	21.5	1172	6	AX175176 Sequence
C 37	92.2	18.9	1167	6	AY061283 Drosophil
C 38	92.2	18.9	2242	3	AY061283 Drosophil
C 39	92.2	18.9	133090	2	AC017815 Drosophil
C 40	92.2	18.9	189904	3	AC093045 Drosophil
C 41	92.2	18.9	197901	3	AC009523 Drosophil
C 42	92.2	18.9	267547	3	AE003623 Drosophil
C 43	86.6	17.7	185872	2	AC068810 Mus muscu
C 44	74.8	15.3	1104	6	AX175168 Sequence
C 45	74.8	15.3	1664	3	AY051529 Drosophil

ALIGNMENTS

RESULT 1
AC011179/c
LOCUS AC011179 153472 bp DNA linear PRI 27-NOV-2001
DEFINITION Homo sapiens, clone RP11-11N24, complete sequence.
ACCESSION AC011179
VERSION AC011179.9 GI:17048255
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 153472)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-11N24
JOURNAL Unpublished

TITLE Direct Submission
JOURNAL Submitted (01-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jan 10, 2002 this sequence version replaced gi:17149773.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8338
Center clone name: 705_O_24

FEATURES

source Location/Qualifiers
1..165720
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="Rp11-705024"
/clone_lib="RPC1-11 Human Male BAC"
complement(1..155)
/rpt_family="MER4B"
494..523
/rpt_family="AT_rich"
578..599
/rpt_family="AT_rich"
3379..3551
/rpt_family="MLT1J"
3652..3693
/rpt_family="AT_rich"
complement(393..4312)
/rpt_family="AluY"
5721..5878
/rpt_family="L1MC4a"
5919..6553
/rpt_family="L1MB3"
6564..6592
/rpt_family="AT_rich"
6603..7052
/rpt_family="MLT1C"
7646..7674
/rpt_family="AT_rich"
7978..8107
/rpt_family="MIR"
8524..8589
complement(8592..8777)
/rpt_family="MSTB1"
complement(8820..8873)
/rpt_family="MSTB1"
8875..8923
/rpt_family="L2"
complement(10625..11047)
/rpt_family="MLT1C"
12127..12148
/rpt_family="AT_rich"
13527..13813
/rpt_family="AluSx"
13817..13841
/rpt_family="AT_rich"
complement(14152..14273)
/rpt_family="FRAM"
complement(15286..16222)
/rpt_family="L2"
17886..18069
/rpt_family="LTR33"
19251..19312
/rpt_family="(CA)n"
complement(19955..20075)

repeat_region /rpt_family="L2"
20867..20917
/rpt_family="AT_rich"
21497..21786
/rpt_family="AluSc"
complement(22069..22292)
/rpt_family="L2"
complement(22427..22907)
/rpt_family="L2"
23178..23211
/rpt_family="(TTG)n"
complement(23667..23791)
/rpt_family="L2"
23960..24038
/rpt_family="MLT1K"
24046..24142
/rpt_family="MLT1K"
complement(24144..24231)
/rpt_family="L2"
24669..24693
/rpt_family="(TTTG)n"
complement(26144..26438)
/rpt_family="AluSc"
27613..27662
/rpt_family="(TAGA)n"
complement(28078..28381)
/rpt_family="AluSg"
28947..29054
/rpt_family="L1MB2"
29502..29920
/rpt_family="L1MB2"
31112..31285
/rpt_family="MIR"
complement(31377..31556)
/rpt_family="MIR"
complement(31689..31816)
/rpt_family="MERSA"
31817..31999
/rpt_family="MERSA"
complement(32000..32051)
/rpt_family="MERSA"
complement(32137..32286)
/rpt_family="FRAM"
32589..33275
/rpt_family="MER67C"
33688..33712
/rpt_family="AT_rich"
33827..34528
/rpt_family="MER50"
34559..34759
/rpt_family="MER50-int"
34735..35321
/rpt_family="MER50-int"
35363..35551
/rpt_family="MER67C"
35697..36152
/rpt_family="MER50-int"
complement(36153..36461)
/rpt_family="AluSx"
36462..37932
/rpt_family="MER50-int"
complement(39608..39756)
/rpt_family="L2"
complement(39968..40488)
/rpt_family="L2"
41542..41605

Query Match 97.8%; Score 478; DB 9; Length 165720;
Best Local Similarity 99.8%; Pred. No. 5,9e-101;
Matches 489; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 CATCTAAAGACGTGATGAAGTTGATTGCAATGCTAGTCAATATAATACCAACACA 60
|||||

Db	26488	CATCTAAAAGACTGATGAAGTTGATTGCAAAATGCTAGTCATCATATAATACCAGAACACA	26547
QY	61	GGTGTCACCTGACAAACTCTATCAAAAGATGAAATTTCTTGTGGATTATGACAGACCT	120
Db	26548	GGTGTCACCTGACAAACTCTATCAAAAGATGAAATTTCTTGTGGATTATGACAGACCT	26607
QY	121	CAAAATCTAG-AAAAAAGATCAGACGCATCAGAGATACATGGGCCAGGGTTGCAATAA	179
Db	26508	CAAAATCTGAAAAAAGATCAGACGCATCAGAGATACATGGGCCAGGGTTGCAATAA	26667
QY	180	AGCGTTGTTTATGAGCTCAAAAGAAAAATAAGACTTCTCTACTGTGGGATACACACAA	239
Db	26568	AGCGTTGTTTATGAGCTCAAAAGAAAAATAAGACTTCTCTACTGTGGGATACACACAA	26727
QY	240	AGAGACAGAAACCAACTGTCCTGGAAAAAGTTTAAAGCTTTCTATATGCTCATGACCA	299
Db	26728	AGAGACAGAAACCAACTGTCCTGGAAAAAGTTTAAAGCTTTCTATATGCTCATGACCA	26787
QY	300	TTATCTGGGAATACATGSGATTGGTTTCATGAAGACAGATGATATATATATATCAC	359
Db	26788	TTATCTGGGAATACATGSGATTGGTTTCATGAAGACAGATGATATATATATATCAC	26847
QY	360	ATTGGACAACCTTGAATGCTTCTCACAAACTATTAACCCCTGATGAATCCACTTACTTTGG	419
Db	26848	ATTGGACAACCTTGAATGCTTCTCACAAACTATTAACCCCTGATGAATCCACTTACTTTGG	26907
QY	420	GAAGAATTTAAGCACTCGCAGAAAACAGCACTACATGACTGGAGAGCAGGATATGACT	479
Db	26908	GAAGAATTTAAGCACTCGCAGAAAACAGCACTACATGACTGGAGAGCAGGATATGACT	26967
QY*	480	GAGCAAAGAA	489
Db	26968	GAGCAAAGAA	26977

RESULT 3	HSAL132443	930 bp	mRNA	linear	PRI 27-JUL-2001
LOCUS	HSAL132443				
DEFINITION	Homo sapiens mRNA for beta-1,3-galactosyltransferase b3gal-T8.				
ACCESSION	U71254.3				
VERSION	U71254.3.1				
KEYWORDS	GI:15028815				
SOURCE	b3gal-T8; b3gal-T8 gene; beta-1,3-galactosyltransferase.				

FEATURES
source

gene

CDS

```

/product="beta-1,3-galactosyltransferase b3Gal-T8"
/protein_id="CAC45046.1"
/db_xref="GI:15028816"
/translation="MASKSWLNF LTF LCGSAIGFLICSLQSFLLICEKVDPTQPNVLHN
DPHARHSDDGQNHLEGMNFADNSOHHKQDIDTAENLYQKVRILCKVWMTGPNLEK
KAKHVATAGRCNKVLFMSSENKQFPVAGLTKEGRDQLYWKTKAFQYVHEHYLE
DADWFLKADDTTVILNDRLWLLSKYDPEEPIYFGRFRFYPKQGYMSGAGVYLSK
ALKRFVDAFKTDKCTHSSSDIALGRCKEIMNVEAGDSRDITKGTETFPFVPEHHLI
KGYLPRTFWYNTNYPYPPVEVSLLEILLCCQYLD"
312 a 161 c 192 g 265 t
BASE COUNT

```

Query Match 62.8%; Score 307; DB 9; Length 930;
Best Local Similarity 78.3%; Pred. No. 5.2e-61;

QY
1 CATCTAAAAAGACTGATGAAGTTGATTGC AAAA TCCTACCTCATCATAAA TCCCA

[illegible]

Db
172 CATCTAGAGGACAAATGAACTTCAATGCAGATTCTAGCCAACATAAAGATGAGAACACA 231

QY
61 GGTGTCACAGACAATCTATCAAAAGATCAAATAATTCTTTTCCCTCCTCGTGCTGG

120

DB 232 GACATTGCTGAAACCTCTATCAGAAAGTTAGAAATTCTTTGCTGGTTATGACCGCCCT 291

QY 121 CAAAATCTAGAAAAAAGATTCAGACCGCATCACACACTACATCCGCCGCTTCTTTCTC

[illegible]

DB 292 CAAACCTAGAGAAAAGGCCAACACGTCAAAGCTACTTGGGCCACGCTGTAACTAA 351

[illegible][illegible]

DB
352 GTGTTGTTATGAGTTCAGAAGAAATAAAGACTTCCCTGCTGTGGGACTGAAACCAA 411

QY
241 GAAGACAGAAACCCTGTCCTCCAAAATACTTAAACCTTTTCCTCCTAT

[illegible]

Db
412 GAAGGCAGAGATCAACTATACTGGAACAATTAAAGCTTTTCAGTATGTTTCATGAACAT 471

QY
301 TATCTGGAATACATGGATTGGTTCATCAAGCACCATCATACTGTACTCGTTA

AUTHORS	Jensen, M.A. and Bennett, E.P.
TITLE	Cloning of a new member of the beta-1,3-galactosyltransferase family; pSGal-18
REFERENCE	Unpublished
AUTHORS	Jensen, M.A. and Bennett, E.P.
TITLE	Cloning of a new member of the beta-1,3-galactosyltransferase family; pSGal-18
JOURNAL	Submitted

```
1. .930
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```
/map="7p14-p13"
```

1. .930
(continued)

1: .930

/gene="b3Gal-T8"

/codon_start=1

```
BASE COUNT      363 a   195 c   219 g   315 t
ORIGIN
Query Match      62.8%; Score 307; DB 6; Length 1092;
Best Local Similarity 78.3%; Pred. No. 5.le-61;
Matches 383; Conservative 0; Mismatches 100; Indels 6; Gaps 1;

Qy 1 CATCTAAAGACATGATGAGTTCATTCGAAATGCTAGTCATCATATAATACCAAGACACA 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 172 CATCTAGAGACAAATGAATTCATCAATCCAGATTTCTAGCCCAACATAAAGATGAGAACA 231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 GGTGCTACTGACAAACTCTCAAAAGATGAAATTTCTTCTGCTGGATTATGCACAGGACCT 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 232 GACATTCCTGANAACCTCTATCAGAAAGTTAGAATTTCTTCTGCTGGTTATGACCGGCCT 291
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 121 CAAATCTAGAAAAAAGATCAGACGATCAGAGATACATGCGCCGAGGTTGCAATAAA 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 292 CAAAACCTAGAGAAAAAGGCCAACACGTCAAAAGCTACTTTGGGCCAGCGTTGTAACAAA 351
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 181 GCGTTGTTATGAGCTCAAAAGAAATTAAGACTTCTCTACTGTGGATTATGCATGACCAT 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 352 GTGTTGTTATGAGTTCAGAAAGAAATTAAGACTTCTCTCTGCTGGACTGAAACACAAA 411
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 241 GAAGACAGAACCAACTGCTCTGGAATAGTAAAGCTTTTCTPATATGCTCATGACCAT 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 412 GAAGCAGAGATCAACTATCTGGAAGAAATTAAGACTTCTCTGCTGGACTGAAACACAAA 471
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 301 TATCTGAATACATGATGTTGTTCAATGAAAGCAGATGATATATATATATATATACACA 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 472 TATTTAGAAGATGCTGATGTTTGAAGACAGATGATGA-----CACGTATGTCATA 525
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 361 TTGGACAACCTTGAATGGCTTCTCACAACATATAACCTGATGAATCCACTTACTTTGGG 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 526 CTAGACAATTTAGCTGCTCTTTCAAAATACGACCTGAGAACCCATTTACTTTGGG 585
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 421 AAAAGATTTAAGCCTATGTAAGACGGGTACATGATGAGTGAGGAGCAGGATATGTACTG 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 586 AGAAGATTTAAGCCTATGTAAGACGGGTACATGATGAGTGAGGAGCAGGATATGTACTG 645
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 481 AGCAAGAA 489
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 646 AGCAAGAA 654

RESULT 5
HSA243256      1092 bp      mRNA      linear      PRI 04-JAN-2002
LOCUS
DEFINITION      Homo sapiens mRNA for beta-1,3-Galactosyltransferase.
ACCESSION      AJ243256
VERSION      b3Gal-T gene; beta-1,3-galactosyltransferase.
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1
                Jensen,M.P.A.
                Cloning and expression of a novel beta-1,3-Galactosyltransferase
                Unpublished
REFERENCE      2 (bases 1 to 1092)
                Jensen,M.P.A.
                Direct Submission
                Submitted (23-JUN-1999) Jensen M.P.A., School of Dentistry,
                University of Copenhagen, Noerre Allé 20, DK-2200 Copenhagen N,
                DENMARK
FEATURES
    source      Location/Qualifiers
                1..1092
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="7"
                /map="7p14-p13"
                1..1092
                /gene="b3Gal-T"

gene
```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1092)
AUTHORS Bennett,E.P.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-2000) Bennett E.P., Histo. lab., 24 5.33, School
of Dentistry, Noerre Alle' 20, 2200N, Copenhagen, DENMARK

FEATURES

Location/Qualifiers
source
1..1092
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1092
/gene="core-1-beta 1,3-galactosyltransferase"
1..1092
/gene="core-1-beta 1,3-galactosyltransferase"
/function="Core-1 synthase"
/codon_start=1
/evidence=experimental
/product="beta 1,3-galactosyltransferase"
/protein_id="CAC82373.1"
/db_xref="GI:17384686"
translation="MASKSWLNFLFLCGSAIGFLCSQLFSILLGEKVDTPQNVLHN
DPHARSDNCONHLEGOMFNADSSQHKDENTDAENLYQKVRILCWMTGPONLEK
KAKHVKAWAQCRNKLFLMSSEENKDPFVGLKTEGRDQLYWKTIKAFQVHEHYLE
DADWFLKADDDTYVILNRLWLLSKYDPEEPIYGRKPKYVKQYMSGGAGYVLSKE
ALKRFVAFKTDKCTHSSIEDLALGRCEIMNVEAGSDRTIGKETPHFPVPEHLLI
KGYLPRTFWYNNYYPVEGPGCCSDLAVSFHYVDSTMYELEYLVYHLRPGYLYR
YQPTLPERILKEISOANKNEDTKVLGNP"
BASE COUNT 363 a 195 c 219 g 315 t
ORIGIN

Query Match 62.8%; Score 307; DB 9; Length 1092;
Best Local Similarity 78.3%; Pred. No. 5.1e-61;
Matches 383; Conservative 0; Mismatches 100; Indels 6; Gaps 1;
QY 1 CATCTAAAAAGACTGATGAAGTTGATTCGAATGCTAGTCATCATATAATACCAGAACACA 60
DB 172 CATCTAGAAGGACAAATGAATTCATCGAATTCAGCCACATATAAGATGAGAACACA 231
QY 61 GGTGTCACGACAACTCTATCAAAAGATGAAATTCCTTGTGCTGATTTATGACAGACCT 120
DB 232 GACATGTGAAACCTCTATCAGAAAGTTAGAATTCCTTGTGCTGATTTATGACAGACCT 291
QY 121 CAAATCTAGAAAAAGATCAGCGCATCAGAGATACATGCGCCCGGTTGCAATAAA 180
DB 292 CAAACCTTAGAAAAAGCCAAACACGCTCAAAGCTACTTGGCCCGGCTGTGAACAAA 351
QY 181 GCGTTGTTTATGAGCTCAAAAGAAATAAGACTTCTCTACTGTGCGGATTACACACAAA 240
DB 352 GTGTTGTTTATGAGTTAGAGAAATAAGACTTCTCTCTGCTGCGGACTGAAACCCAAA 411
QY 241 GAAGACAGAACCACTGCTCTGGAATAAGTAAAGCTTTTCTATATGCTCATGACCAT 300
DB 412 GAAGCAGAGATCACTACTACTGGAACAAATTAAGCTTTTTCAGTATGTTATCAATCA 471
QY 301 TATCTGGAATACATGATGTTTCAATGAAGCAGATGATATATATATATATATACA 360
DB 472 TATTAGAAAGTGTGATGTTTGTGAAAGCAGATGATGA-----CACGTATGTCATA 525
QY 361 TTGGCAACTTGAATGCTCTCAACAACTATACCCCTGATGAATCCACTTACTTTGGG 420
DB 526 CTAGCAATTTGAGTGCTCTCTCAAAATACGACCTGAAAGACCCCTTACTTTGGG 585
QY 421 AAAAGATTAAAGCACTGCAGAAAAACAGCAGCTACATGACTGGAGGAGCAGGATATGACTG 480
DB 586 AGAAGATTAAAGCCCTTATGTAAGACGAGGCTACATGAGTGGAGGAGCAGGATATGACTA 645
QY 481 AGCAAGAA 489
DB 646 AGCAAGAA 654

RESULT 7
AXI75160

LOCUS AXI75160
DEFINITION Sequence 2 from Patent WO0144478.
ACCESSION AXI75160
VERSION AXI75160.1 GI:14598562
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1794)
AUTHORS Canfield,W.M., Cummings,R.D. and Ju,T.
TITLE Core lb3-galactosyl transferases and methods of use thereof
JOURNAL Patent: WO 0144478-A 2 21-JUN-2001;
The Board of Regents of The University of Oklahoma (US)

FEATURES

Location/Qualifiers
source
1..1794
/organism="Homo sapiens"
/db_xref="taxon:9606"
63..1154
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC43870.1"
/db_xref="GI:14598563"
translation="MASKSWLNFLFLCGSAIGFLCSQLFSILLGEKVDTPQNVLHN
DPHARSDNCONHLEGOMFNADSSQHKDENTDAENLYQKVRILCWMTGPONLEK
KAKHVKAWAQCRNKLFLMSSEENKDPFVGLKTEGRDQLYWKTIKAFQVHEHYLE
DADWFLKADDDTYVILNRLWLLSKYDPEEPIYGRKPKYVKQYMSGGAGYVLSKE
ALKRFVAFKTDKCTHSSIEDLALGRCEIMNVEAGSDRTIGKETPHFPVPEHLLI
KGYLPRTFWYNNYYPVEGPGCCSDLAVSFHYVDSTMYELEYLVYHLRPGYLYR
YQPTLPERILKEISOANKNEDTKVLGNP"
BASE COUNT 624 a 300 c 334 g 536 t
ORIGIN

Query Match 62.8%; Score 307; DB 6; Length 1794;
Best Local Similarity 78.3%; Pred. No. 4.7e-61;
Matches 383; Conservative 0; Mismatches 100; Indels 6; Gaps 1;

QY 1 CATCTAAAAAGACTGATGAAGTTGATTCGAATGCTAGTCATCATATAATACCAGAACACA 60
DB 234 CATCTAGAAGGACAAATGAATTCATCGAATTCAGCCACATATAAGATGAGAACACA 293
QY 61 GGTGTCACGACAACTCTATCAAAAGATGAAATTCCTTGTGCTGATTTATGACAGACCT 120
DB 294 GACATGTGAAACCTCTATCAGAAAGTTAGAATTCCTTGTGCTGATTTATGACAGACCT 353
QY 121 CAAATCTAGAAAAAGATCAGCGCATCAGAGATACATGCGCCCGGTTGCAATAAA 180
DB 354 CAAACCTTAGAAAAAGCCAAACACGCTCAAAGCTACTTGGCCCGGCTGTGAACAAA 413
QY 181 GCGTTGTTTATGAGCTCAAAAGAAATAAGACTTCTCTACTGTGCGGATTACACACAAA 240
DB 414 GTGTTGTTTATGAGTTAGAGAAATAAGACTTCTCTCTGCTGCGGACTGAAACCCAAA 473
QY 241 GAAGACAGAACCACTGCTCTGGAATAAGTAAAGCTTTTCTATATGCTCATGACCAT 300
DB 474 GAAGCAGAGATCACTACTACTGGAACAAATTAAGCTTTTTCAGTATGTTATCAATCA 533
QY 301 TATCTGGAATACATGATGTTTCAATGAAGCAGATGATATATATATATATATACA 360
DB 534 TATTAGAAAGTGTGATGTTTGTGAAAGCAGATGATGA-----CACGTATGTCATA 587
QY 361 TTGGCAACTTGAATGCTCTCAACAACTATAACCCCTGATGAATCCACTTACTTTGGG 420
DB 588 CTAGCAATTTGAGTGCTCTCTCAAAATACGACCTGAAAGACCCCTTACTTTGGG 647
QY 421 AAAAGATTAAAGCACTGCAGAAAAACAGCAGCTACATGACTGGAGGAGCAGGATATGACTG 480
DB 648 AGAAGATTAAAGCCCTTATGTAAGACGAGGCTACATGAGTGGAGGAGCAGGATATGACTA 707
QY 481 AGCAAGAA 489
DB 708 AGCAAGAA 716


```
QY 181 GCCTGCTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACAAA 240
Db 352 GTGTTATTTATGAGTTTCAGAAAGAAATAAAGACTTCTCTACTGTGGGCTGAAACAAA 411
QY 241 GAAGCAGAGAAACCAACTGCTCGTGGAAATAGTTAAAGCTTTCTATATGCTCATGACCAT 300
Db 412 GAAGCAGAGAAACCAACTGCTCGTGGAAACAAATTAAGCGTTTCAGTACGTACATGACCAT 471
QY 301 TATCTGGAATACATGAGTGGTTGTTTCAAGAGCAGATGATATATGATATATATATACCA 360
Db 472 TACTTAGAAGATGCTGACTGGTTTATGAAAGCAGACGATGA-----CACCTATGTCATA 525
QY 361 TTGGCAACACTGAAATGGCTTCTCAAAACTATAACCCCTGATGAATCCACTTACTTTGGG 420
Db 526 CTGGACAATCTGAGATGGCTTCTATCAAAAGTATAACCCCTGAACAGCCCATTTACTTTGGG 585
QY 421 ARAAGATTAAAGCAGTGCAGAAACAGGACTACATGCTGGAGGACGAGATATGATCTG 480
Db 586 AGAAGATTAAAGCCCTATGTGAAGCAGGATACATGAGTGGAGGACGAGATATGTCCTA 645
QY 481 AGCAAGAA 489
Db 646 AGCAAGAA 654

RESULT 10
AF157963
LOCUS
DEFINITION
Rattus norvegicus corel UDP-galactose:N-acetylglactosamine-alpha-R
beta 1,3-galactosyltransferase (Cigalt1) mRNA, complete cds.
AF157963
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1405)
Ju,T., Brewer,K., D'Souza,A., Cummings,R.D. and Canfield,W.M.
Cloning and Expression of Human Core 1 beta
1,3-Galactosyltransferase
J. Biol. Chem. 277 (1), 178-186 (2002)
11677243
2 (bases 1 to 1405)
Ju,T., Cummings,R.D. and Canfield,W.M.
Cloning and characterization of corel
UDP-galactose:N-acetylglactosamine-alpha-R beta
1,3-galactosyltransferase (Corel beta3-Gal-T) cDNA sequence
Unpublished
3 (bases 1 to 1405)
Ju,T., Cummings,R.D. and Canfield,W.M.
Direct Submission
Submitted (10-JUN-1999) Department of Medicine, The University of
Oklahoma, Health Science Center, 975 N.E. 10th Street, Biomedical
Research Center, Room 411, Oklahoma City, OK 73104, USA
Location/Qualifiers
1. 1405
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1. 1405
/gene="Cigalt1"
154. 1245
/gene="Cigalt1"
/EC_number="2.4.1.122"
/function="responsible for the synthesis of corel
structure of O-glycans"
/notes="corel beta3-Gal-T"
/codon_start=1
/product="corel
UDP-galactose:N-acetylglactosamine-alpha-R beta
1,3-galactosyltransferase"
/protein_id="AAF81983.1"
```

```
/db_xref="GI:8927166"
/translation="MASKSWLNFLFLCGSAIGFFLCQLLNILLQEQADVPQNMHLHN
DPHARISDSGHNHLKGOMDFNADSSQHDNDVAENLYQKVLKQWMTSPONLEK
KAKHKVATWAORCNKVLFMSSSENKDFPTVGLTEKGBOLYWKTKATQYVHDVLE
DADWENKADDDTYVILDNLRLWLLSKYNPEQPIYFGRFRPYKQGYMSGAGYVLSKE
ALURFVDAEKTEKCTHSSSDIALGRCHMEILKVEAGDSRDPTGKTFTHFFVPEHHLI
KGYLPKTFWYNNYYPVPEGCGCCSDIAVSPHYSDTSTMYELEYLVYHLRPGYLYR
YQPALPENILKEINQVKKEDTKIKLGNP"
misc_feature
166. 249
/gene="Cigalt1"
/notes="Region: non-cleavable transmembrane signal"
/evidence=not_experimental
BASE COUNT 445 a 288 c 300 g 372 t
ORIGIN

Query Match 61.5%; Score 300.6; DB 10; Length 1405;
Best Local Similarity 77.5%; Pred. No. 1.5e-59;
Matches 379; Conservative 0; Mismatches 104; Indels 6; Gaps 1;

QY 1 CATCTAAAAGACTGATGAAGTTGATTGCAAAATGCTAGTCATCATATAATACCAACACA 60
Db 325 CACCTCAAAGGACAGATGGACTTCAATGCAGATTCTAGCCAACATAAAGATGAGAACACA 384
QY 61 GGTGTCACCTGACAAACTCTATCAAAAGATGAAAATTTCTTTGCTGGATTATGACAGACCT 120
Db 385 GACGTTCTCTGAGAACCTCTATCAGAAAGTTAAAGTTCTTTTGTGGTTATGACAAGCCT 444
QY 121 CAAAATCTAGAAAAAAGATCAGACGCATCAGAGATACATGGCCGAGGTTTCAATAAA 180
Db 445 CAAAATCTAGAAAAAAGGCCAACACAGTCAAAGTACATGGGCCGAGCCTTTGTAATAA 504
QY 181 GCGTTGTTATGAGCTCAAAAGAAAATAAAGACTTCTCTACTGTGGATTATACACACAAA 240
Db 505 GTCTTATTATGAGTTTCAGAGAATAAAGACTTCCCTACTGTGGGCTGGAACACAAA 564
QY 241 GAAGCAGAAACCAACTGCTCGTGGAAAATAGTTTAAAGCTTTTCTATATGCTCATGACCAT 300
Db 565 GAAGCAGAGAGCAACTGTACTGGAAAACAATTTAAAGCGTTTCAGTACGTACATGACCAT 624
QY 301 TATCTGGAATACATGAGTGGTTTCATGAAGCAGATGATCATATATGATATATATACACA 360
Db 625 TACTTAGAAGATGCTGCTGGTTTATGAAAGCAGACGATGA-----CACCTATGTCATA 678
QY 361 TTGGCAACACTTGAATGGCTTCTCAACAAACTATAACCCCTCATGAATCCACTTACTTTGGG 420
Db 679 CTGGACAATCTGAGATGGCTTCTCTCAAGTATATAACCCCTGAACAGCCCTTACTTTGGG 738
QY 421 AAAAGATTTAAGCACTGCAGAAAACAGGACTACATGACTGGAGGACGAGGATATGTAATG 480
Db 739 AGAAGATTTAAGCCCTATGTGAAGCAGGATACATGAGTGGAGGACGAGGATATGTCCTA 798
QY 481 AGCAAGAA 489
Db 799 AGCAAGAA 807

RESULT 11
AX175162
LOCUS
DEFINITION
Sequence 4 from Patent WO0144478.
AX175162
ACCESSION
AX175162
VERSION
AX175162.1 GI:14598564
KEYWORDS
Norway rat.
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1440)
Canfield,W.M., Cummings,R.D. and Ju,T.
Core lb3-galactosyl transferases and Methods of use thereof
Patent: WO 0144478-A 4 21-JUN-2001;
The Board of Regents of The University of Oklahoma (US)
```


[illegible]

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 1708)
Strausberg, R.
Direct Submission
Submitted (13-MAR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaps@email.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 45 Row: j Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
source

Location/Qualifiers

1..1708
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="IMAGE:5040878"
/tissue_type="kidney, normal. 5 month old male mouse."
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
<1..903
/codon_start=1
/product="Similar to core 1
UDP-galactose-4-epimerase
1,3-galactosyltransferase"
/protein_id="AAH25899.1"
/db_xref="GI:19484215"
/translation="NENASDQKHNDIDVAENLYQKYLKLVWVTPQNLEKAKHY
KATWQRQKLVFMSSENDFTVLGKTEGREQLYKTKIKAFQYVHDHLEADWF
MKADDDTYIVDNLKLLSKYNPEQPIYFGRREKPYKQYMGSGAGYVLSKEALRRF
VNAFKTEKTHSSIEDLALGRCEIINVEAGDSRDTIGKTEPHFVPEHHLIKGLP
KTFWYNNYYPPIEGPGCCSDIAVSFHYVDGTTMVELEYLVVHLRYPYGLYRYQPAL
PENLKEINQVRKEDTKIKLNP"

BASE COUNT 590 a 315 c 315 g 488 t
ORIGIN

Query Match 57.5%; Score 281; DB 10; Length 1708;
Best Local Similarity 76.4%; Pred. No. 4.9e-55;
Matches 360; Conservative 0; Mismatches 105; Indels 6; Gaps 1;

QY 19 AAGTTGATTCGAATGCTAGTCATCAATAATACAGAACACAGGTGTCTACTGACAACTC 78
DB 1 AACTTCAATCGATTCCAGCCCAACATAAAGATCAGAACATAGACGTGTCTGAGAACCTC 60
QY 79 TATCAAAAGATGAAATCTTTCTGCTGATATGACAGACCTCAAAATCTAGAAAAAAG 138
DB 61 TATCAGAAAGTTAAATCTTTGTGGTTATGACAGATCTCTCAAAATCTAGAAAAAAG 120
QY 139 ATCAGAGCATCAGACATACATGGCCCGAGGTTCGAATAAAGCGTTGTTTATGAGCTCA 198
DB 121 GCCAAACATGTCAAAGCTACGTGGCCCGACCGTTGTAATAAAGTGTTATTATGAGTTCG 180
QY 199 AAGAAATAAAGACTTCTCTACTGTGGGATTACACACCAAGAAAGACAAACCAACTG 258
DB 181 GAAGAAATAAAGACTTCTCTACTGTGGGATTGAAACCAAGAAAGACGAGAGCAACTA 240

QY 259 TCCTGGAAATAGTTAAAGCTTTTCTATATGCTCATGACCATTTATCTGGAATACATGGAT 318
DB 241 TATTGGAAACAATTAAGCTTTCCAGTATGTACATGACCATTTATTTAGAAGATGCTGAC 300
QY 319 TGGTTTCATGAAAGCAGATGAT 378
DB 301 TGGTTTATGAAAGCAGATGACGA-----CACATACGTCTATTGTGGACAACCTGAGATGG 354
QY 379 CTTCACAAACTATACCTGATGATCCACTTACTTTGGGAAAGATTTAAGCACTGC 438
DB 355 CTTCATCAAGATATACCTGATGATCCACTTACTTTGGGAAAGATTTAAGCACTGC 414
QY 439 AGAAAAACAGGACTACATGACTGTGGAGGAGCAGGATATGCTACTGAGCAAGAA 489
DB 415 GTGAAGCAGGATACATGACCGGAGGAGCGGCTATGCTCTTAAGCAAGAA 465

Search completed: June 28, 2003, 00:22:44
Job time : 1593.73 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 22:29:21 ; Search time 192.373 Seconds
(without alignments)
5724.434 Million cell updates/sec

Title: US-10-029-359A-3

Perfect score: 489

Sequence: 1 catctaaagactgatgaa.....gatatgtactgagcaagaa 489

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	489	100.0	489	24	Human DNA encoding
2	460.6	94.2	795	22	Human breast cancer
3	397.6	81.3	513	22	Human breast cancer
4	307	62.8	1794	22	Human core 1 beta3
5	307	62.8	1800	21	Nucleotide sequenc
6	305.4	62.5	1560	24	Human DNA encoding
7	300.6	61.5	1440	22	Rat core 1 beta3-g
8	292.6	59.8	1469	22	Mouse core 1 beta3
9	278.6	57.0	1424	23	DNA encoding novel

C 10	278	56.9	1975	22	AA136608	Probe #5294 used t
C 11	278	56.9	1975	24	ABS05361	Human genome-deriv
C 12	277.4	56.7	668	22	AA149726	Probe #18412 used
C 13	277.4	56.7	668	24	ABS17860	Human genome-deriv
C 14	168	34.4	505	24	ABK88407	Human DNA encoding
C 15	105	21.5	1172	22	AAF90510	Caenorhabditis ele
C 16	92.2	18.9	1167	22	AAF90512	Drosophila core 1
C 17	92.2	18.9	1432	23	ABL20061	Drosophila melanog
C 18	92.2	18.9	1519	23	ABL20059	Drosophila melanog
C 19	92.2	18.9	1727	23	ABL15335	Drosophila melanog
C 20	92.2	18.9	4660	23	ABL20060	Drosophila melanog
C 21	92.2	18.9	9262	23	ABL15334	Drosophila melanog
C 22	92.2	18.9	9262	23	ABL20058	Drosophila melanog
C 23	90.6	18.5	2043	23	ABL23893	Drosophila melanog
C 24	79.6	16.3	1158	23	ABL05467	Drosophila melanog
C 25	74.8	15.3	1104	22	AAF90511	Drosophila core 1
C 26	74.8	15.3	1474	23	ABL05129	Drosophila melanog
C 27	74.8	15.3	3641	23	ABL05128	Drosophila melanog
C 28	60.6	12.4	4553	23	ABL23892	Drosophila melanog
C 29	54.4	11.1	4058	23	ABL12932	Drosophila melanog
C 30	54	11.0	1310	23	ABL12933	Drosophila melanog
C 31	51.4	10.5	3442	23	ABL05466	Drosophila melanog
C 32	49.2	10.1	1672	23	ABL20549	Drosophila melanog
C 33	49.2	10.1	3442	23	ABL05466	Drosophila melanog
C 34	49.2	10.1	3814	23	ABL20548	Drosophila melanog
C 35	41.8	8.5	7244	22	AA546726	Tumour suppressor
C 36	41.6	8.5	1827	24	ABN92451	Staphylococcus epi
C 37	41.6	8.5	3003	22	AAH54803	S. epidermidis gen
C 38	38.8	7.9	60	24	ABN42559	Human spliced tran
C 39	38.6	7.9	640861	24	ABA92787	Buchnera sp. genom
C 40	37.6	7.7	3714	22	AAH54411	S. epidermidis gen
C 41	36.8	7.5	6095	22	AA546309	Tumour suppressor
C 42	36.8	7.5	6095	24	ABL70149	Chemically treated
C 43	36.8	7.5	6095	24	ABL32360	Human immune syste
C 44	36.8	7.5	6095	24	ABL34474	Human metastasis a
C 45	36.8	7.5	33053	24	ABQ67006	Human angiogenesis

ALIGNMENTS

RESULT 1	ABK88408	ABK88408 standard; DNA; 489 BP.
ID	ABK88408	standard; DNA; 489 BP.
AC	ABK88408	
XX	07-OCT-2002	(first entry)
DT	Human DNA encoding partial protein cluster II protein #2.	
DE	Human; ds; gene; protein cluster II; obesity; diabetes mellitus;	
KW	central nervous system disorder; metabolic disease.	
XX	Homo sapiens.	
OS	Key	Location/Qualifiers
XX	CDS	55...489
FT		/*tag= a
FT		/product= "Protein cluster II protein #2"
FT		/note= "No start or stop codon shown"
XX	WO200251864-A1.	
PN	04-JUL-2002.	
XX	14-DEC-2001; 2001WO-SF02786.	
PD	22-DEC-2000; 2000SE-0004828.	
XX	(PHAA) PHARMACIA AB.	
XX	Attersand A;	

XX WPI: 2002-575368/61.
DR P-PSDB; ABG30852.
XX
PT New protein cluster II nucleic acids and polypeptides, useful in
PT diagnosing metabolic diseases such as obesity and diabetes, and in
PT identifying agents for treating such diseases
PS
PS Claim 1; Page 21; 28pp; English.
XX
XX The invention relates to a new isolated nucleic acid comprising:
CC (a) a fully defined sequence, encoding a protein cluster II protein,
CC appearing as ABK88407 and ABK88408; (b) a sequence capable of hybridising
CC under stringent hybridisation conditions to a nucleotide sequence
CC complementary to the polypeptide coding region of a nucleic acid in (a);
CC or (c) a sequence which is degenerate as a result of the genetic code to
CC a nucleotide sequence in (a) or (b). Also included are a isolated
CC an polypeptide encoded by the nucleic acid, a vector harbouring the
CC nucleic acid, a replicable expression vector, which carries and is
CC capable of mediating the expression of the nucleic acid a cultured host
CC cell harbouring the expression vector, producing a polypeptide by
CC culturing the cell, where the polypeptide is produced, and recovering the
CC polypeptide and identifying an agent capable of modulating the nucleic
CC acid by providing a cell comprising the nucleic acid molecule, contacting
CC the cell with a candidate agent, and monitoring the cell for an effect
CC that is not present in the absence of the candidate agent. The protein
CC cluster II nucleic acid and polypeptide is useful in the diagnosis of
CC metabolic diseases such as obesity and diabetes, and central nervous
CC system disorders and in the identification of agents for treating these
CC diseases. The nucleic acids may be used as hybridisation probes, for
CC chromosome and gene mapping, in polymerase chain reaction (PCR)
CC technologies, in the production of sense and antisense nucleic acids, and
CC in screening for new therapeutic molecules. The present sequence
CC encodes a protein cluster II protein of the invention.
XX
SQ Sequence 489 BP; 187 A; 86 C; 95 G; 121 T; 0 other;

Query Match 100.0%; Score 489; DB 24; Length 489;
Best Local Similarity 100.0%; Pred. No. 9e-124;
Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATCTAAAGAACTGATGAAGTTGATTCGAAATGCTAGTCATCATATAATACCAGAACACA 60
DB 1 CATCTAAAGAACTGATGAAGTTGATTCGAAATGCTAGTCATCATATAATACCAGAACACA 60

QY 61 GGTGTCACCTGACAACTCTATCAAGATGAAATTCCTTCTGCTGATATGACAGACCT 120
DB 61 GGTGTCACCTGACAACTCTATCAAGATGAAATTCCTTCTGCTGATATGACAGACCT 120

QY 121 CAAATCTAGAAAAAGATCAGACGATCAGAGATACATGCGCCAGGTTGCAATAAA 180
DB 121 CAAATCTAGAAAAAGATCAGACGATCAGAGATACATGCGCCAGGTTGCAATAAA 180

QY 181 GCGTTGTTTATGAGTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACCAA 240
DB 181 GCGTTGTTTATGAGTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACCAA 240

QY 241 GAAGACAGAAACCACTGCTCGGAAATAGTAAAGCTTTCTATATGCTCATGACCAT 300
DB 241 GAAGACAGAAACCACTGCTCGGAAATAGTAAAGCTTTCTATATGCTCATGACCAT 300

QY 301 TATCTGGAATACATGGATGTTGTTATGAAAGCAGATGATATATATATATATATATAT 360
DB 301 TATCTGGAATACATGGATGTTGTTATGAAAGCAGATGATATATATATATATATATAT 360

QY 361 TTGGACAACTTGAATGGCTTCTCACAACATATAACCTGTAGTAATCCACTTACTTTGG 420
DB 361 TTGGACAACTTGAATGGCTTCTCACAACATATAACCTGTAGTAATCCACTTACTTTGG 420

QY 421 AAAAGATTTAAGCACTGCAGAAAAACAGACTACATGCTGGAGGAGCAGGATATCTACTG 480
DB 421 AAAAGATTTAAGCACTGCAGAAAAACAGACTACATGCTGGAGGAGCAGGATATCTACTG 480

QY 481 AGCAAGAA 489
DB 481 AGCAAGAA 489

RESULT 2
AAL24100/c
ID AAL24100 standard; cDNA; 795 BP.
XX
XX AAL24100;
AC
XX
XX 07-DEC-2001 (first entry)
DT
XX
XX Human breast cancer expressed polynucleotide 16557.
DE
XX
XX Human; breast cancer; cell marker; cytostatic; ss.
KW
XX
XX Homo sapiens.
OS
XX
XX WO200151628-A2.
PN
XX
XX 19-JUL-2001.
PD
XX
XX 10-JAN-2001; 2001WO-US00798.
PF
XX
XX 14-JAN-2000; 2000US-0176077.
PR
XX
XX 14-MAR-2000; 2000US-0189167.
PR
XX
XX 24-MAR-2000; 2000US-0192099.
PR
XX
XX 29-MAR-2000; 2000US-0193480.
PR
XX
XX 15-MAY-2000; 2000US-0205230.
PR
XX
XX 09-JUN-2000; 2000US-0211315.
PR
XX
XX 25-JUL-2000; 2000US-0220534.
XX
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA
XX
XX Lillie J, Xu Y, Wang Y, Steinmann K;
PI
XX
XX WPI: 2001-451856/48.
DR
XX
XX New peptide useful as a marker for the diagnosis of breast cancer
PT
XX
XX Claim 1; Page 3031-3032; 3695pp; English.
PS
XX
XX The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 795 BP; 202 A; 171 C; 151 G; 271 T; 0 other;

Query Match 94.2%; Score 460.6; DB 22; Length 795;
Best Local Similarity 99.0%; Pred. No. 5.9e-116;
Matches 474; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 CATCTAAAGAACTGATGAAGTTGATTCGAAATGCTAGTCATCATATAATACCAGAACACA 60
DB 481 CATCTAAAGAACTGATGAAGTTGATTCGAAATGCTAGTCATCATATAATACCAGAACACA 422

QY 61 GGTGTCACCTGACAACTCTATCAAGATGAAATTCCTTCTGCTGATATGACAGACCT 120
DB 421 GGTGTCACCTGACAACTCTATCAAGATGAAATTCCTTCTGCTGATATGACAGACCT 362

QY 121 CAAATCTAG-AAAAAAGATCAGACGATCAGAGATACATGCGCCAGGTTGCAATAA 179
DB 361 CAAATCTAGAAAAAAGATCAGACGATCAGAGATACATGCGCCAGGTTGCAATAA 302

QY 180 ASCGTTGTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACCAA 239
DB 180 ASCGTTGTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACCAA 239

Db 301 ACGCTTGTATGAGCTCAAAAGAAAATAAGACTTCTATATCTGTGGATTACACACCAA 242
QY 240 AGAAGACAGAACCAACTGTCTCTGGAATAAGTAAAGCTTTCTATATGCTCATGACCA 299
Db 241 AGAAGACAGAACCAACTGTCTCTGGAATAAGTAAAGCTTTCTATATGCTCATGACCA 182
QY 300 TTATCTGGAAATACATGATGTTCTCATGAAGCAGATGATGATATATATATATAC 359
Db 181 TTATCTGGAAATACATGATGTTCTCATGAAGCAGATGATGATATATATATATAC 122
QY 360 ATTGACAACTTGAATGGCTTCTCACAACCTATACCTGATGAATCCACTTACTTTGG 419
Db 121 ATTGACAACTTGAATGGCTTCTCACAACCTATACCTGATGAATCCACTTACTTTGG 62
QY 420 GAAAGATTTAAGCACTGACAGAAAACAGGACTACATGATGAGGAGCAGGATATGTAC 478
Db 61 GAAAGATTTAAGCACTGACAGAAAACAGGACTACATGATGAGGAGCAGGATATGTAC 3

RESULT 3

AA115254/c
ID AAL15254 standard; cDNA; 513 BP.
XX AC AAL15254;
XX XX
DT 07-DEC-2001 (first entry)
XX Human breast cancer expressed polynucleotide 7711.
DE Human; breast cancer; cell marker; cytostatic; ss.
XX Homo sapiens.
OS
KW
XW
PN WO200151628-A2.
XX PD
XX 19-JUL-2001.
XX PF 10-JAN-2001; 2001WO-US00798.
XX PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX XX

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Lillie J, Xu Y, Wang Y, Steinmann K;

WPI; 2001-451856/48.

PS New peptide useful as a marker for the diagnosis of breast cancer -
XX Claim 1; Page 1386; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
XX activity.

SQ Sequence 513 BP; 133 A; 90 C; 87 G; 180 T; 23 other;

Query Match 81.38; Score 397.6; DB 22; Length 513;
Best Local Similarity 93.84; Pred. No. 8e-99;
Matches 435; Conservative 0; Mismatches 24; Indels 5; Gaps 3;

QY 1 CATCTAAAAAGACTGATG-AAAGTGTATTGCAAAATGCTAGTCATCA--TAAATACACAGAC 57
Db 464 CATCTAAAAAGANNTNTNTAGTTGTAATGCTAGTATCATCAATAAAATTCACAGAAC 405
QY 58 ACAGTGTCTACTGACAAACTCTATCAAAAGATGAAAATTTCTTGTGCTGATTTATGACAGGA 117
Db 404 ACAGTGTCTACTGCAAACTCTATCAAAANATGAAAATTTCTTGTGCTGATTTATGNCAGSN 345
QY 118 CCTCAAAATCT--AGAAAAAAGATCAGACGCATCAGAGATACATGGCCCCAGGGTTGCA 175
Db 344 CCTCAAAATCTTAGAANAAAAAGATCAGACGNATCAGAGATACATGGCCCCAGGGTTGCA 285
QY 176 ATAAAGCCTTGTATGAGCTCAAAAGAAAAATAAGACTTCTCTACTGTGGGATTACACA 235
Db 284 ATAAAGCCTTGTATGAGCTCAAAAGAAAAATAAGACTTCTCTACTGTGGGNTNACACA 225
QY 236 CCAAGAAGACAGAAACCACTGTCTCTGGAATAAGTAAAGCTTTTCTATATGCTCATG 295
Db 224 CCAANAAGACAGAAACCACTGTCTCTGGANATAGTTANAGCNTTCTATATGCTCNTG 165
QY 296 ACCATTATCTGGAATACATGATGGTTTCATGAAGCAGAGATGATGATATATATATATA 355
Db 164 ACCATTATCTGGAATACATGATGGTTTCATGAAGCAGAGATGATGATATATATATATA 105
QY 356 TCACATTGGACAACTTGAATGGCTTCTCAAAACTATAACCCCTGATGAATCCACTTACT 415
Db 104 TCACATTGGACAACTTGAATGGCTTCTCNAACTATAACCCCTGATGAATCCACTTACT 45
QY 416 TTGGGAAAAGATTTAAGCACTGACAGAAACAGGACTACATGACT 459
Db 44 TTGGGAAAAGATTTAAGCCTGCAAGAAACAGGACTACATGACT 1

RESULT 4

AAF90507
ID AAF90507 standard; cDNA; 1794 BP.
XX AC AAF90507;

XX 22-AUG-2001 (first entry)

XX Human core 1 beta3-galactosyl transferase cDNA.

XX Core 1 beta3-galactosyl transferase; human; O-glycosylation;
KW galactosylation; glycosulfopeptide; Tn syndrome; IgA nephropathy;
KW diagnosis; therapy; ss.
XX Homo sapiens.

OS

XX Key Location/Qualifiers

FT CDS 63..1154

FT /*tag= a

FT /note= "a polynucleotide comprising this
FT coding sequence is also claimed in Claim
FT 1(A)"

XX WO200144478-A2.

PN 21-JUN-2001.

XX 14-DEC-2000; 2000WO-US33945.

XX 15-DEC-1999; 99US-0461321.

XX 15-DEC-1999; 99US-0464035.

XX (UYOK-) UNIV OKLAHOMA.

XX Canfield WM, Cummings RD, Ju T;

XX WPI; 2001-398157/42.

XX P-PSDB; AAB82456.

XX Novel purified core 1 beta3 galactosyl transferases of human, rat,

Canfield WM, Cummings RD, Ju T;
WPI: 2001-398157/42.
P-PSDB; AAB82458.

Novel purified core 1 beta3 galactosyl transferases of human, rat, mouse, Drosophila melanogaster, Caenorhabditis elegans, useful for galactosylating N-acetyl galactosamine linked to O-linking amino acid on protein

Claim 1(A); Page 64; 74pp; English.

The present sequence is that of mouse cDNA for core 1 beta3-galactosyltransferase (b3GTase, see AAB82458). The mouse core 1 b3GTase has 89% identity and 94% similarity to the human enzyme (see AAB82456). The invention provides human, rat, mouse, Drosophila melanogaster and Caenorhabditis elegans b3GTases and nucleic acids encoding them. The polypeptides exhibit a wide range of homologies. The polynucleotides can be used to transform or transfect host cells for producing substantially pure forms of the enzyme, or for use in an expression system for core 1 O-linked glycosylation of proteins or peptides produced within the expression system. The expressed enzymes can be used to galactosylate, via a beta3 linkage, an N-acetylgalactosamine linked to a serine, threonine or other O-linking amino acid on peptides or proteins. Core 1 b3GTases are also useful in the synthesis of glycosulfopeptides which can function as inhibitors of P-selectin:PSGL-1 interactions. Other potential uses include diagnostic tests for the rare Tn-syndrome or IgA nephropathy, and therapy of these disorders.

Sequence 1469 BP; 467 A; 296 C; 313 G; 393 T; 0 other;

Query Match 59.8%; Score 292.6; DB 22; Length 1469;
Best Local Similarity 76.5%; Pred. No. 5.1e-70;
Matches 374; Conservative 0; Mismatches 109; Indels 6; Gaps 1;

QY 1 CATCTAAAAGACTGATGAAGTTGATTCGAAATGCTAGTCATCATATAATACCAGAACACA 60
Db 351 CACCTCAAGGACAGATGAATCTCAATGTCAGATTCACGCCAACATAAAGATGAGAACATA 410
QY 61 GGTGTCTACTCACAACCTCTATCAAAAGATGAAATCTTTGTCGTGATATGACAGGACCT 120
Db 411 GAGTTGCTCAGAACCTCTATCAAGAGTTAAATCTTTGTGGGTTATGACAGAGTCCT 470
QY 121 CAAATCTAGAAAAAAGATGACAGCGCATCAGAGATACATGGGCCAGGGTTGCAATAAA 180
Db 471 CAAAATCTAGAAAAAAGGCGCAACATGTCAAAGCTACGTGGGCCAGCGTTGTAATAAA 530
QY 181 GCCTTTGTTATGAGCTCAAAAGAAAATAAGACTTCTCTACTGTGGGATACACACAAA 240
Db 531 GTGTTATTTATGAGTTCGGAAGAAAATCAAGACTTCCTACTGTGGGATGAAAACCAA 590
QY 241 GAAGACAGAACCAACTGCTCGGAAAATAGTTAAAGCTTTCTATATGCTCATGACCAT 300
Db 591 GAAGGCAGAGACAATATTCGAAACAAATTAAGCTTTCCAGTAIGTACATGACCAT 650
QY 301 TATCTGGAATACATGATTGGTTTTCATGAAAGCAGATGATATATGTATATATATACCA 360
Db 651 TATTTAGAAGTCTGACTGGTTTATGAAAGCAGATGACGA-----CACATACGTCATT 704
QY 361 TTGGAACAATTGAAATGGCTTCTCACAAACTATAACCTTGATGAATCCACTTACTTTGGG 420
Db 705 GTGGACAACCTGAGATGGGTTCTTATCAAAAGTATAACCTTGAACAGCCATTTACTTTGG 764
QY 421 AAAAGATTTTAAGCACTGCGAAAAACAGGACTACATGACTGGAGSAGCAGATGATGCTG 480
Db 765 CGAAGATTTAAGCCCTATGTGAACGAGGATACATGACGGGAGCGGCTATGTCCCTA 824
QY 481 AGCAAGAA 489
Db 825 AGCAAGAA 833

Db 248 GACATTCCTGAAAGGCTCGAGGAGAAAGTTAGAAATCTTTGCTGGTTATGACCGGCC 307
Qy 120 TC-AAAATCTAGAAAAAAGATCAGACGATCAGAGATACATGGGCCCGAGGTTGCAATA 178
Db 308 TCAAAACCTAGAAAAAAGCCAAACACGTCAAAGCTACTTGGGCCCGAGGTTGTARCA 367
Qy 179 AAGCGTTGTTTATGAGCTCAAAAGAAAAATAAAGACTTCTCTACTGTGGGATTACACACCA 238
Db 368 AAGTGTGTTTATGAGTTCAGAGCAAAATAAAGACTTCCCTGCTGTGGACTGAAACCA 427
Qy 239 AAGAGACACAGAACCAACTGCTCGGAAATAGTAAAGCTTTTCTATATGCTCATGACC 298
Db 428 AAGAGGACAGAGTCAACTATCTCTGGAACAAATTAAGCTTTTCACTGTTTCATGAAC 487
Qy 299 ATTATCTGGAATACATCGATTGGTTTCATGAAGAGAGATGATATATGATATATATCA 358
Db 488 ATATTATATAGCTGATGGTTTGTGAAGACAGATGATGA-----CACGTATGTCA 541
Qy 359 CATTTGACAACTTGAATGGCTTCTCAAACTATAACCTGATGAATCCACTTTACTTTG 418
Db 542 TACTAGACAATTTGAGGTGGCTTCTTCAAAATACGACCTGGAAGAACCCATTTCATTG 601
Qy 419 GGAAGAGATTTAAGCACTGCAGAAACAGGAGACTACATGACTGGAGGAGGATATGTAC 478
Db 602 GGAGAGATTTAAGCCTTATGTAAAGCAGGGCTACATGAGTGGAGGAGGATATGTAC 661
Qy 479 TGAGCAAAAGAA 489
Db 662 TAAGCAAGAA 672

RESULT 10
AAI36608/c
ID AAI36608 standard; DNA; 1975 BP.
AC AAI36608;
XX
XX
DT 17-OCT-2001 (first entry)
XX
XX
DE Probe #5294 used to measure gene expression in human placenta sample.
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX
DR WPI; 2001-488897/53.
XX
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX
PS Claim 25; SEQ ID No 5294; 654pp; English.
XX
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 1975 BP; 627 A; 395 C; 347 G; 606 T; 0 other;
Query Match 56.9%; Score 278; DB 22; Length 1975;
Best Local Similarity 78.4%; Pred. No. 5.4e-66;
Matches 348; Conservative 0; Mismatches 90; Indels 6; Gaps 1;
Qy 46 AAATACAGACACAGCAGTGTCTACTGACAAAATCTATCAAAAGATGAAATCTTTGCTGG 105
Db 938 ATAGATCAGACACACAGATTTGCTGAAACCTCTATCAGAAAAGTTAGAAATCTTTGCTGG 879
Qy 106 ATTATGACAGACCTCAAAATCTAGAAAAGATCAGACGATCAGAGATACATGAGGCC 165
Db 878 GTTATGACCGCCCTCAAAACCTAGAGAAAAGGCCAAACACGTCAAGCTACTTTGGGCC 819
Qy 166 CAGGTTTGCATTAAGCGTTGTTTATGAGCTCAAAAGAAAATAAAGACTTTCTCTACTGTG 225
Db 818 CAGCGTTGTAACAAAGTGTTTATGAGTTTCAGAGAAAATAAAGACTTCCCTGCTGTG 759
Qy 226 GGATTACACACCAAGAGACAGAACCAACTGCTCTGGAATAAGTAAAGCTTTTCTA 285
Db 758 GGACTGAAAACCAAGAGGACGAGATCAACTACTGGAACCAATTAAGCTTTTTCAG 699
Qy 286 TATGCTCATGACCAATTATCTGGAATACATGATGTTGTTGATGAAGCAGATGATATA 345
Db 698 TATGTTTCATGAACATTTATAGAGATGCTGATGTTGTTTGAAGCAGATGATGA---- 643
Qy 346 TGTATATATATCACAATTTGGACAACTTGAATGCTTCTCACAACTATAACCTGATGAA 405
Db 642 --CAGGTATGCTACTAGACAATTTGAGGTGGCTTCTTCAAAATACGACCTGGAAGAA 585
Qy 406 TCCACTTACTTTGGGAAAAGATTTAAGCACTGCAAGAAAACAGCACTACATGCTGGAGGA 465
Db 584 CCCATTTACTTTGGGAGAAGATTTAAGCCTTATGTAAGCAGGGCTACATGAGTGGAGGA 525
Qy 466 GCAGGATATGCTACTGAGCAAGAA 489
Db 524 GCAGGATATGCTACTAAGCAAGAA 501
RESULT 11
ABS05361/c
ID ABS05361 standard; DNA; 1975 BP.
XX
XX
AC ABS05361;
XX
XX
DT 19-AUG-2002 (first entry)
XX
XX
DE Human genome-derived single exon probe from lung SEQ ID No 5352.
XX
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200186003-A2.
XX
XX
PD 15-NOV-2001.
XX
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.

```

PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples -
XX
PS Claim 1; SEQ ID No 5352; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a single exon
CC probe of the invention.
CC
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1975 BP; 627 A; 395 C; 347 G; 606 T; 0 other;

Query Match          56.9%; Score 278; DB 24; Length 1975;
Best Local Similarity 78.4%; Pred. No. 5.4e-66;
Matches 348; Conservative 0; Mismatches 90; Indels 6; Gaps 1;

QY 46 AAATACAGAACACAGTGTCACTGACAACTCTATCAAAAGATGAAATCTTCTCTGG 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 938 ATAGATGAGACACACATGCTGGAACCTCTATCAGAAATTTAGATTTCTCTGG 879
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 106 ATTATGACGAGGACCTCAAAATCTAGAAAAAAGATCAGACGCGCATCAGATCGGCG 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 878 GTTATGACGGCGCCTCAAAACCTAGAGAAAAAGCCCAACACGCTCAAAAGCTACTTGGCC 819
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 CAGGGTTGCAATAAAGCGTTGTTATGAGCTCAAAAGAAAAAAGAGCTTCTACTGTG 225
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 818 CAGGTTGTACAAAGTGTGTTTATGAGTTCAGAAAGAAATAAAGACTTCCCTGCTGTG 759
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 226 GGATTACACACCACAAAGAGACAGAAACCAACTGTCTCTGGAAATAGTTAAAGCTTTTCTA 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 758 GGACTGAAACCAACAAAGAGCAGAGATCAACTACTACTGGAACAATTAAGCTTTTTCAG 699
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 286 TATGCTCATGACCATTTATCTGGATACATGGATGTTGTTCAATGAAGCAGATGATGATATA 345
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 698 TATGTTTCATGAACATTTATTTAGAGATGCTGATTTGTTTGAAGCAGATGATGA---- 643
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 346 TGTATATATATCATCAATTTGGACAACTTGAATGGCTTCTCACAACATATAACCTTGATGAA 405
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 642 --CACGTATGTCATCTACTAGACATTTGAGGTGCTTCTTCAAAATACGACCTGGAAGA 585
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 406 TCACATTTACTTTGGGAAAAGATTTAAGCACTGCGAGAAACAGCAGCTACATGACTGGAGGA 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 584 CCCATTTACTTTGGGAGAAGATTTAAGCCCTATGTAAGCAGGCGCTACATGAGTGGAGGA 525
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 466 GCAGGATATGCTACTGACCAAGAA 489
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 524 GCAGGATATGCTACTAAGCAAGAA 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 12

AAI49726/C

ID AAI49726 standard; DNA; 668 BP.

XX AAI49726;

AC AAI49726;

XX AAI49726;

DT 17-OCT-2001 (first entry)

XX AAI49726;

DE Probe #18412 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

XX Homo sapiens.

OS Homo sapiens.

XX Homo sapiens.

PN WQ200157272-A2.

XX WQ200157272-A2.

PR 09-AUG-2001.

XX 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US006663.

XX 30-JAN-2001; 2001WO-US006663.

PF 04-FEB-2000; 2000US-0180312.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

XX 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

XX 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

XX 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

XX 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

XX 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

DR WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -

XX Human genome-derived single exon nucleic acid probes useful for

PS Claim 25; SEQ ID No 18412; 654pp; English.

XX Claim 25; SEQ ID No 18412; 654pp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).

XX The present invention relates to single exon nucleic acid probes (SENP).

CC The present sequence is one such probe. The probes are useful for

XX The present sequence is one such probe. The probes are useful for

CC producing a microarray for predicting, measuring and displaying gene

XX producing a microarray for predicting, measuring and displaying gene

CC expression in samples derived from human placenta. The probes are useful

XX expression in samples derived from human placenta. The probes are useful

SQ for antenatal diagnosis of human genetic disorders.

SQ for antenatal diagnosis of human genetic disorders.

SQ Sequence 668 BP; 178 A; 144 C; 115 G; 231 T; 0 other;

SQ Sequence 668 BP; 178 A; 144 C; 115 G; 231 T; 0 other;

Query Match 56.7%; Score 277.4; DB 22; Length 668;

Query Match 56.7%; Score 277.4; DB 22; Length 668;

QY 293 ATGACCAATTATCGGAATACATGATGGTTTCATGAAAGCAGATGATGATATATATAT 352
 Db IIII IIII I III I IIII IIII IIII IIII IIII IIII I
 425 ATGAACATATTATTAGAAGATGCTGATGGTTTTTGAAGCAGATGATGA-----CACGT 372
 QY 353 ATATCACAATTGGCAACACTTGAATGGCTTCTCACAACATATATACCCCTGATGATCCACTT 412
 Db IIII I IIII IIII IIII IIII IIII IIII IIII IIII IIII
 371 ATGTCATACACTAGACAATTTGAGGTGGCTTCTTCAAAATACGACCCCTGAAGAACCCTATTT 312
 QY 413 ACTTTGGGAAAGATTTAAGCACTGCGCAAGAACACAGGACTACATGACTGGAGGACGAGAT 472
 Db IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 311 ACTTTGGGAGAGATTTAAGCCCTTATGTAAGCAGGCTACATGATGGAGGACGAGAT 252
 QY 473 ATGTAAGTGAAGAA 489
 Db IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII
 251 ATGTAAGTGAAGAA 235

RESULT 14

ID ABK88407 standard; DNA; 505 BP.

AC ABK88407;

DT 07-OCT-2002 (first entry)

XX Human DNA encoding partial protein cluster II protein #1.

XX Human; ds; gene; protein cluster II; obesity; diabetes mellitus;
 XX central nervous system disorder; metabolic disease.

OS Homo sapiens.

XX Key Location/Qualifiers
 FT CDS 21..497

FT /*tag= a
 FT /product= "Protein cluster II protein #1"
 FT /note= "No stop codon shown"
 FT /transl_except= (pos:297,299,aa:Xaa)
 FT /note= "Xaa= unknown"

PN WO200251864-A1.

XX 04-JUL-2002.

XX 14-DEC-2001; 2001WO-SE02786.

XX 22-DEC-2000; 2000SE-0004828.

XX (PHAA) PHARMACIA AB.

XX Attersand A;

XX WPI; 2002-575368/61.

DR P-PSDB; ABG30851.

XX New protein cluster II nucleic acids and polypeptides, useful in
 PT diagnosing metabolic diseases such as obesity and diabetes, and in
 PT identifying agents for treating such diseases

PS Claim 1; Page 20-21; 28pp; English.

XX The invention relates to a new isolated nucleic acid comprising:
 CC (a) a fully defined sequence, encoding a protein cluster II protein,
 CC appearing as ABK88407 and ABK88408; (b) a sequence capable of hybridising
 CC under stringent hybridisation conditions to a nucleotide sequence
 CC complementary to the polypeptide coding region of a nucleic acid in (a);
 CC or (c) a sequence which is degenerate as a result of the genetic code to
 CC a nucleotide sequence in (a) or (b). Also included are a isolated
 CC an polypeptide encoded by the nucleic acid, a vector harbouring the
 CC nucleic acid, a replicable expression vector, which carries and is
 CC capable of mediating the expression of the nucleic acid a cultured host
 CC cell harbouring the expression vector, producing a polypeptide by
 CC culturing the cell, where the polypeptide is produced, and recovering the

CC polypeptide and identifying an agent capable of modulating the nucleic
 CC acid by providing a cell comprising the nucleic acid molecule, contacting
 CC the cell with a candidate agent, and monitoring the cell for an effect
 CC that is not present in the absence of the candidate agent. The protein
 CC cluster II nucleic acid and polypeptide is useful in the diagnosis of
 CC metabolic diseases such as obesity and diabetes, and central nervous
 CC system disorders and in the identification of agents for treating these
 CC diseases. The nucleic acids may be used as hybridisation probes, for
 CC chromosome and gene mapping, in polymerase chain reaction (PCR)
 CC technologies, in the production of sense and antisense nucleic acids, and
 CC in screening for new therapeutic molecules. The present sequence
 CC encodes a protein cluster II protein of the invention.

XX Sequence 505 BP; 166 A; 95 C; 108 G; 135 T; 1 other;

Query Match 34.4%; Score 168; DB 24; Length 505;

Best Local Similarity 74.7%; Pred. No. 3.7e-36; Indels 0; Gaps 0;
 Matches 210; Conservative 0; Mismatches 71;

QY 1 CATCTAAAAAGACTGATGAAGTTTCATTCGCAATGCTAGTCATCATATAATACCAACACACA 60
 Db IIII I IIII IIII IIII IIII IIII IIII IIII IIII IIII

216 CATCTAGGAGGACAAATGAACCTTCAATGCAGATCTAGCCCAACCTAAGATGAGACACA 275
 Db IIII I IIII IIII IIII IIII IIII IIII IIII IIII IIII

QY 61 GGTGTCACCTCACAACTCTATCAAAAGATGAAATCTTTGCTGGATATGACAGGACCT 120
 Db I IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

276 GAAATCGCTGAAACCTCTATNAGCAAGTTAAAAATCTTTGCTGGTTATGACAGGCTCT 335
 Db I IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

QY 121 CAAAATCTAGAAAAAGATCAGACGCATCAGAGATACATGGCCCCAGGCTTGCATAAA 180
 Db I IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

336 CAAAACCTACAGAAAAAGGCCAACATGTCAAAGCTACATGGGCCAGCGTTGTCTAAAA 395
 Db I IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

QY 181 GCGTGTGTTTATGAGCTCAAAAGAAAAATAAAGACTTCTCTACTGTGGATTTACACACCAA 240
 Db I IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

396 GTATTTTATGAGTTCAGAGAAAAATAAGACTTCGCTGCTGTGGATTTGAAACCCAAA 455
 Db I IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

QY 241 GAAGCAGAAAAACCACTGCTCTCGAAAAATAGTTAAAGCTTTT 281
 Db I IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII

456 GCAGGCAGAGATGAGCTATATCTGCAAAAAACAATTAACCTTTT 496

RESULT 15

AAF90510

ID AAF90510 standard; cDNA; 1172 BP.

XX AAF90510;

XX 22-AUG-2001 (first entry)

XX Caenorhabditis elegans core 1 beta3-galactosyl transferase cDNA.

XX Core 1 beta3-galactosyl transferase; nematode; O-glycosylation;
 KW galactosylation; glycosulfopeptide; Tn syndrome; IgA nephropathy;
 KW diagnosis; therapy; ss.

XX Caenorhabditis elegans.

OS Key Location/Qualifiers

PH CDS 1..1170

FT /*tag= a

FT /note= "a polynucleotide comprising this
 FT coding sequence is also claimed in Claim
 FT 1(A)"

XX WO200144478-A2.

XX 21-JUN-2001.

XX 14-DEC-2000; 2000WO-US33945.

XX 15-DEC-1999; 99US-0461321.

PR 15-DEC-1999; 99US-0464035.

XX (UOK-) UNIV OKLAHOMA.

```
XX Canfield WM, Cummings RD, Ju T;
PI WPI: 2001-398157/42.
DR P-PSDB; AAB82459.
XX
PT Novel purified core 1 beta3 galactosyl transferases of human, rat,
PT mouse, Drosophila melanogaster, Caenorhabditis elegans, useful for
PT galactosylating N-acetyl galactosamine linked to O-linking amino acid
PT on protein
XX
XX Claim 1(A); Page 66; 74pp; English.
XX
CC The present sequence is that of Caenorhabditis elegans cDNA for
CC core 1 beta3-galactosyltransferase (b3GTase, see AAB82459). The
CC nematode b3GTase has 41% identity and 58% similarity to the human
CC enzyme (see AAB82456). The invention provides human, rat, mouse,
CC Drosophila melanogaster and Caenorhabditis elegans b3GTases and
CC nucleic acids encoding them. The polypeptides exhibit a wide
CC range of homologies. The polynucleotides can be used to transform
CC or transfect host cells for producing substantially pure forms of
CC the enzyme, or for use in an expression system for core 1 O-linked
CC glycosylation of proteins or peptides produced within the
CC expression system. The expressed enzymes can be used to
CC galactosylate, via a beta3 linkage, an N-acetylgalactosamine linked
CC to a serine, threonine or other O-linking amino acid on peptides or
CC proteins. Core 1 b3GTases are also useful in the synthesis of
CC glycosulfopeptides which can function as inhibitors of
CC P-selectin:PSGL-1 interactions. Other potential uses may include
CC diagnostic tests for the rare Tn-syndrome or IgA nephropathy, and
CC therapy of these disorders.
XX
SQ Sequence 1172 BP; 321 A; 290 C; 289 G; 272 T; 0 other;
```

Query Match 21.58; Score 105; DB 22; Length 1172;
Best Local Similarity 55.2%; Pred. No. 7.6e-19;
Matches 229; Conservative 0; Mismatches 180; Indels 6; Gaps 1;

```
QY 68 CTGACAACTCTATCAAAAGATGAAATCTTTGCTGGATTATGACAGGACCTCAAAATC 127
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 290 CCGACGAAGTTGCGAGAAGGTTGCGCTCTTCTGTGGATTCTCACCGGAAACAGATC 349
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 128 TAGAAAAAAGATCAGACGCATCAGAGATACATGGGCCAGGGTTGCAATAAAGCGTTGT 187
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 350 ACGATAAACGCGGAAACACGCTCAAGCCACCTGGGCCAAGCGGTGTAATAGTACGTAT 409
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 188 TTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGTGGGATTACACACCACAAAGACAC 247
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 410 TCATGTCTCAGAAGAGGATGCGAGAACTCCCGACCATCAACCTAAATGTCTCTGAAGGCA 469
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 248 GAAACCAACTGTCTCTGAAATAGTTAAAGCTTTTCTATATGCTCATGACCATTTATCTGG 307
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 470 GAGATTACTTGTGGGCAAAACAAAGGAGCATTCAAATACATTTATGACCATCACCTGA 529
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 308 AATACATGGATTGGTTCAATGAAAGCAGATGATATATGATATATATATACATTTGGACA 367
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 530 ACGACTACGACTGGTTCCTGAAAGCCGACGACGCA-----TACCTATGTGGTGTGGA 583
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 368 ACTTGAATGGCTTCACAAACTATAACCTCATGATCCACTTACTTTGGGAAAGAT 427
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 584 ATCTCGGGTTTCATGCTATTGGCTCATTCACCGGATGAGCCAATTCACCTTTGGATGCAAGT 643
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 428 TTAAGCACTGCAGAAACAGGATACATGACTGGAGGAGGAGGATATGTACTGAG 482
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 644 TTAAGCCATTTCACACAGGGAGGATATCATAGTGGTGGAGCTGGATATGTGCTCAG 698
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: June 27, 2003, 23:39:44
Job time : 193.373 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 23:09:36 ; Search time 41.5774 Seconds
(without alignments)
3606.885 Million cell updates/sec

Title: US-10-029-359A-3
Perfect score: 489
Sequence: 1 catctaaagactgatgaa.....gatattactgagcaagaa 489

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	45.8	9.4	7218	1	US-08-232-463-14
2	41.6	8.5	1827	4	US-09-134-001C-1914
3	36	7.4	668	4	US-09-605-785-367
4	36	7.4	668	4	US-09-439-313-367
5	36	7.4	668	4	US-09-062-451-293
6	36	7.4	668	4	US-09-353-616A-367
C 7	35	7.2	1569	3	US-08-821-984-9
C 8	35	7.2	1569	4	US-09-328-749-9
C 9	35	7.2	6895	1	US-08-570-311-9
C 10	35	7.2	6895	2	US-08-353-485-9
11	34.8	7.1	4967	4	US-09-177-325-12
12	34.8	7.1	4967	4	US-09-411-812A-12
13	34.8	7.1	4967	4	US-09-590-113-12
C 14	34.2	7.0	272	4	US-09-221-017B-617
15	33.2	6.8	1175	4	US-09-378-088A-77
16	33.2	6.8	1410	4	US-09-378-088A-81
C 17	33	6.7	80595	4	US-09-078-294-3
C 18	32.6	6.7	782	4	US-09-221-017B-522
C 19	32.6	6.7	7685	4	US-09-221-017B-1092
C 20	32.4	6.6	99500	4	US-09-798-096-10
21	32.2	6.6	424	4	US-09-308-003-9
22	32.2	6.6	972	4	US-09-134-001C-952
23	32.2	6.6	1071	4	US-09-308-003-10
24	32.2	6.6	2685	4	US-09-221-017B-616
25	32.2	6.6	10754	2	US-08-966-958-1
26	32.2	6.6	10754	2	US-09-215-817-1
27	32.2	6.6	10754	4	US-09-342-353-1

28	32	6.5	2223	1	US-08-257-073-4	Sequence 4, Appli
29	31.8	6.5	2760	4	US-09-198-484-1	Sequence 1, Appli
30	31.4	6.4	5053	4	US-08-961-527-187	Sequence 187, App
31	31	6.3	1092	4	US-09-221-017B-136	Sequence 136, App
32	31	6.3	1386	4	US-09-221-017B-986	Sequence 986, App
33	31	6.3	4325	2	US-08-888-497-21	Sequence 21, Appl
34	31	6.3	4325	4	US-09-362-230-21	Sequence 21, Appl
35	31	6.3	4325	5	PCT-US94-07926-21	Sequence 21, Appl
C 36	31	6.3	6173	4	US-09-221-017B-898	Sequence 898, App
C 37	30.8	6.3	1665	4	US-09-134-001C-1565	Sequence 1565, Ap
C 38	30.8	6.3	1972	1	US-08-463-048-1	Sequence 1, Appli
C 39	30.8	6.3	1972	1	US-08-463-229-1	Sequence 1, Appli
C 40	30.8	6.3	1972	2	US-08-302-891-1	Sequence 1, Appli
41	30.8	6.3	20165	4	US-09-609-816-7	Sequence 7, Appli
42	30.8	6.3	51259	3	US-08-781-891-209	Sequence 209, App
43	30.6	6.3	3910	4	US-09-302-620B-91	Sequence 91, Appl
C 44	30.4	6.2	624	4	US-09-318-978-9	Sequence 9, Appli
C 45	30.2	6.2	18073	4	US-09-078-294-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; TOPOLOGY: single
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-fls
US-08-232-463-14

Query Match 9.4%: Score 45.8; DB 1; Length 7218;

Best Local Similarity 7.9%; Pred. No. 0.0011;
Matches 35; Conservative 212; Mismatches 194; Indels 0; Gaps 0;

QY 46 AAATACGAGAACACAGGTGTGCACACAAACTCTATCAAAAAGATGAAAAATTCTTTGGTGG 105
| | | | | | | | | | | | | | | | | | |
Db 1507 AAAAAAGCGCATGTGAGCAGCATGTAATTACCTATCTATCAAGTAGTTAAGAGATAG 1448

QY 106 ATTATGACGAGCCTCAAAATCTAGAAAAAAGATCAGACCATCAGAGATACATGGGCC 165
| | | | | : : : : : : : : : : : : : :
Db 1447 AAGAAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1388

QY 166 CAGGGTTGCAATAAAGCGTGTGTTTATGAGCTCAAAGAATAAAGACTTCTCTACTGTG 225
: : : : : : : : : : : : : : : : : : :
Db 1387 RR 1328

QY 226 GGATTACACACCAGGAAGACAGACAACCAACTGTCTCTGGAAAAAGTAGTTAAAGCTTTTCTA 285
: : : : : : : : : : : : : : : : : : :
Db 1327 RR 1268

QY 286 TATGTCATCACCAATTATCTGGAATACATGGATTGGTTCATCAAGACAGATCATGATATA 345
: : : : : : : : : : : : : : : : : : :
Db 1267 RR 1208

QY 346 TGATATATATACATTGGACAACCTGAAATGGCTTCTCACAAACTATAACCCCTCATCAA 405
: : : : : : : : : : : : : : : : : : :
Db 1207 RR 1148

QY 406 TCACATTACTTTGGGAAAAGATTTAAGCACTGCAGAAAACAGGACTACATGACTGGAGGA 465
: : : : : : : : : : : : : : : : : : :
Db 1147 RR 1088

QY 466 GCAGGATATCTACTGAGCAAA 486
: : : : : : : : : : : : : : : : : : :
Db. 1087 RRRRRRRRRRRRRRRRRRRR 1067

RESULT 2
US-09-134-001C-1914
; Sequence 1914, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1914
; LENGTH: 1827
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1914

```

Qy 369 CTTGAATGGCTTCTCACAACATATAACCGCTGATGAATCCACTTACTTTGGGAAAA 424
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1748 ATGAAAAAGCCTCTGGTGAAAATGCTAAATTCATAAATGAAAAAATATTAAAAAGAA 1803

RESULT 3
US-09-605-785-367
; Sequence 367, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 367
; LENGTH: 668
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-605-785-367

Query Match 7.4%; Score 36; DB 4; Length 668;
Best Local Similarity 60.0%; Pred. No. 0.29;
Matches 60; Conservative 0; Mismatches 40; Indels 0; Gaps 0

Qy 25 ATTGCAATGCTAGTCATCATATAAACACACAGGTGTCACTGACAAACTCTATCAA 84
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 297 ATTACATATCTTAGGAATTCAAAATAACATCCACAGCTTCACCAACTAGTTATATTTA 356

Qy 85 AAGATGAAAATCTTTGTGGATTATGACAGACCTCAA 124
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 357 AAGAGAGAAAATCAATTTTATGCGCATGTATTGAAATCAA 396

```



```

; NAME/KEY: CDS
; LOCATION: 696..5894
US-08-570-311-9

```

Query Match	7.2%	Score 35;	DB 1;	Length 6895;
Best Local Similarity	52.4%	Pred. No. 1.2;		
Matches	77;	Conservative 0;	Mismatches 70;	Indels 0; Gaps 0;
Qy	165	CCAGGGTTCAAATAAAGCGTTGTTTATGAGCTCAAAGAAATAAAGACTTCTCTACTGT	224	
Db	6165	CCATGGAGGAATATATGCTGTGTTTTAACTCAAAATATACAAAATAACTCCCTATTAT	6106	
Qy	225	GGGATTACACCAAGAGAAGACAAACCAACTGTCCTGGAAAATAGTTAAAGCTTTTCT	284	
Db	6105	ACATGNATTACAAACAANAATACACACCAATGCCGTGCCAAGGTCATTTTCAAAG	6046	
Qy	285	ATATGTCATGACCATATTATCTGGAATA	311	
Db	6045	GTCTCTCTTTGACAAGTTTACCGACA	6019	

RESULT 10
US-08-353-485-9/c
: Sequence 9, Application US/08353485
: Patent No. 5830710
: GENERAL INFORMATION:
: APPLICANT: Proquleske-Fox, Ann
: APPLICANT: Tumwasorn, Somying
: APPLICANT: Lepine, Guytaine
: APPLICANT: Han, Naiming
: APPLICANT: Lantz, Marilyn
: APPLICANT: Patti, Joseph
: TITLE OF INVENTION: Cloned Porphyromonas gingivais Genes
: TITLE OF INVENTION: and Probes for the Detection of Periodontal Disease
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ted W. Whitlock
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606

```

; NAME/KEY: CDS
; LOCATION: 696..5894
US-08-353-485-9

```

[illegible]

```

RESULT 11
US-09-177-325-12
; Sequence 12, Application US/09177325B
; Patent No. 6214983
; GENERAL INFORMATION:
; APPLICANT: Tureci, Ozlem
; APPLICANT: Sakin, Ugur
; APPLICANT: Pfreundschuh, Michael
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Encode SCP Proteins
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: LUD 5525
; CURRENT APPLICATION NUMBER: US/09/177,325B
; CURRENT FILING DATE: 1998-10-22
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 12
; LENGTH: 4967
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; US-09-177-325-12

```

	Query Match	7.1%	Score 34.8;	DB 4;	Length 4967;
	Best Local Similarity 48.5%;	Pred. No. 1.2;	Mismatches 102;	Indels 0;	Gaps 0;
	Matches 96; Conservative	0;			
QY	232	CACACAAAAGAACACAGAAACCACACTGCTCGTGAAATAAGTTTAAAGCTTTTCATATATGC	291		
Db	4555	CACAGAACTGATGTCACTATTTCATGCTCTCATGAAGAANAATGCTTAATGTGTGAATCTAGT	4614		
QY	292	CATGACCATTATCTGGAAATACATGGATTGGTTTCATGAAAGCAGATGATATATATGATA	351		
Db	4615	TTTTATCACCATACTTTATCTAATATTATTCTCTGTATATACTGAGGAAATAAGAATA	4674		
QY	352	TATATCACATTGACAACTTGAATGGCTTCTCACAACTATACCCCTGATGAATCCACT	411		
Db	4675	GTCCCTACAAGAGAAAAATATACATGTCACCGAAGCAAGTGTAACCCCTTTATAGGAACCT	4734		
QY	412	TACTTTGGGAAAGATTT	429		
Db	4735	CAAAATAAAAAAAATGT	4752		

```

RESULT 12
US-09-411-812A-12
; Sequence 12, Application US/09411812A
; Patent No. 6261778
; GENERAL INFORMATION:
; APPLICANT: Tureci, Ozlem
; APPLICANT: Sahin, Ugur
; APPLICANT: Pfreundschuh, Michael
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Encode SCP Proteins, A
; TITLE OF INVENTION: Thereof

```

```
; FILE REFERENCE: LUD 5525.1 CIP
; CURRENT APPLICATION NUMBER: US/09/411.812A
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 09/177,325
; PRIOR FILING DATE: 1998-10-22
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 12
; LENGTH: 4967
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-411-812A-12

Query Match 7.1%; Score 34.8; DB 4; Length 4967;
Best Local Similarity 48.5%; Pred. No. 1.2;
Matches 96; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 232 CACACCAAGAGACAGACAGAACCAACTGCTCTGGAAATAGTTAAAGCTTTTCTATATGCT 291
DB 4555 CAGAGAACTGATGCAGTATTCATGCTCATGAAGAAATGCTAATGTGTGAAATCTAGT 4614
QY 292 CATGACCAATTTCTGGAAATACATGATGGTTTCATGAAGCAGATGATGATATATGATA 351
DB 4615 TTTTATCACCATACTTTATCTAATTTATTTCTCTGTATATACTGAGGAATAAGATA 4674
QY 352 TATATCATTGGACAACTTGAATGGCTTCTCACAACACTATAACCCCTGATGAATCCACT 411
DB 4675 GTCTACAAAGAGAAAAATATACATGTCACCGAAGCAAGTGTACCCCTTTATAGGAACCT 4734
QY 412 TACTTTGGGAAAAGATTT 429
DB 4735 CAAATTAATAAAAAATGT 4752

RESULT 13
US-09-590-113-12
; Sequence 12, Application US/09590113
; Patent No. 6306389
; GENERAL INFORMATION:
; APPLICANT: Tureci, Ozlem
; APPLICANT: Pfreundschuh, Michael
; APPLICANT: Sakin, Ugur
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Encode SCP Proteins, And Us
; FILE REFERENCE: LUD 5525
; CURRENT APPLICATION NUMBER: US/09/590.113
; CURRENT FILING DATE: 2000-06-08
; PRIOR APPLICATION NUMBER: US 09/177,325
; PRIOR FILING DATE: 1998-10-22
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 12
; LENGTH: 4967
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-590-113-12

Query Match 7.1%; Score 34.8; DB 4; Length 4967;
Best Local Similarity 48.5%; Pred. No. 1.2;
Matches 96; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 232 CACACCAAGAGACAGACAGAACCAACTGCTCTGGAAATAGTTAAAGCTTTTCTATATGCT 291
DB 4555 CAGAGAACTGATGCAGTATTCATGCTCATGAAGAAATGCTAATGTGTGAAATCTAGT 4614
QY 292 CATGACCAATTTCTGGAAATACATGATGGTTTCATGAAGCAGATGATGATATATGATA 351
DB 4615 TTTTATCACCATACTTTATCTAATTTATTTCTGTATATACTGAGGAATAAGATA 4674
QY 352 TATATCATTGGACAACTTGAATGGCTTCTCACAACACTATAACCCCTGATGAATCCACT 411
DB 4675 GTCTACAAAGAGAAAAATATACATGTCACCGAAGCAAGTGTACCCCTTTATAGGAACCT 4734
QY 412 TACTTTGGGAAAAGATTT 429
```

```
DB 4735 CAAATTAATAAAAAATGT 4752

RESULT 14
US-09-221-017B-617/c
; Sequence 617, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221.017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 617:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...272
US-09-221-017B-617

Query Match 7.0%; Score 34.2; DB 4; Length 272;
Best Local Similarity 58.3%; Pred. No. 0.7;
Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 165 CCAGGGTTGCAATAAAGCGTTGTTTATGAGCTCAAAAGAAATAAAGACTTCTCTACTGT 224
DB 147 CCATGGGAGGAGATATTATGCTGTTTTTAATACTCAAAATATACAAATACTCCCTACTAT 88
QY 225 GGGATTACACACCAAGAGAGACAGAAACCAACTGCTCTCTGGAAA 267
```

Db 87 ACAATCAATTAAACAGCAAAATACACACCAATTCGGTGC AAA 45

RESULT 15

US-09-378-088A-77
; Sequence 77, Application US/09378088A
; Patent No. 6372480
; GENERAL INFORMATION:
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Knuth, Mark
; APPLICANT: Pollard, Michael R.
; APPLICANT: Cardineau, Guy A.
; APPLICANT: Schwab, George E.
; APPLICANT: Michaels, Tracy E.
; APPLICANT: Fiostad Lee, Stacy
; APPLICANT: Burneister, Paula
; APPLICANT: Dojillo, Joanna
; TITLE OF INVENTION: Pesticidal Proteins
; FILE REFERENCE: MA703C2
; CURRENT APPLICATION NUMBER: US/09/378,088A
; CURRENT FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: US 08/633,993
; PRIOR FILING DATE: 1996-04-19
; PRIOR APPLICATION NUMBER: US 08/844,188
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 77
; LENGTH: 1175
; TYPE: DNA
; ORGANISM: Bacillus thuringiensis
US-09-378-088A-77

Query Match 6.8%; Score 33.2; DB 4; Length 1175;
Best Local Similarity 51.5%; Pred. No. 2.1;
Matches 101; Conservative 0; Mismatches 93; Indels 2; Gaps 1;
QY 2 ATCTAAAAGACTGATGAAGTTGATTGCAATGCTAGTCATCATATAATACCAGAACACAG 61
Db 744 ATTTCAAAATTAATGTAGATTCAGGAATGAAGTTTGAGGTACAGAGTAGGAGGTAC 803
QY 62 GTGTCACCTGACAACTCTATCAAAAGATGAAAATTTCTTTGCTGGATTATGACAGGACCTC 121
Db 804 AGAGAGAAATAAAACACAAATTAATGA--AGAATTAAAAGTTGAATATAGCACTGACACC 861
QY 122 AAAATCTAGAAAAAAGATCAGACGCATCAGAGATACATGGGCCCGAGGGTTGCAATAAG 181
Db 862 AAAATAATGAAAAAATATCAAGAACACTCAGAGATAGATAATCCAACTAATCAACAATG 921
QY 182 CGTTGTTTATGAGCTC 197
Db 922 AATCTATAGGATTTC 937

Search completed: June 28, 2003, 00:24:01
Job time : 43.5774 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 23:29:41 ; Search time 126.594 seconds
(without alignments)
5734.045 Million cell updates/sec

Title: US-10-029-359A-3
Perfect score: 489
Sequence: 1 catctaaagactgatgaa.....gatatgactgagcaagaa 489

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	489	100.0	489	9	US-10-029-359A-3
C 2	458.6	93.8	555	9	US-10-198-846-13278
C 3	415.6	85.0	879	9	US-10-198-846-6422
4	307	62.8	1092	9	US-10-235-056-12
5	307	62.8	1794	9	US-10-235-056-2
6	305.4	62.5	1560	9	US-10-029-359A-5
7	300.6	61.5	1092	9	US-10-235-056-13
8	300.6	61.5	1440	9	US-10-235-056-4
9	292.6	59.8	1092	9	US-10-235-056-14
10	292.6	59.8	1469	9	US-10-235-056-6
C 11	278	56.9	1975	10	US-09-864-761-15628
C 12	277.4	56.7	668	10	US-09-864-761-32135
13	105	21.5	1170	9	US-10-235-056-15
14	105	21.5	1172	9	US-10-235-056-8
15	92.2	18.9	1167	9	US-10-235-056-18
16	74.8	15.3	1104	9	US-10-235-056-10
17	46.2	9.4	299	9	US-10-029-359A-1
18	39.2	8.0	409	10	US-09-783-590-8460
19	38.6	7.9	640681	10	US-09-790-988-1

C 20	37	7.6	161	10	US-09-969-373-1319	Sequence 1319, Ap
C 21	36.6	7.5	652	10	US-09-746-801A-38	Sequence 38, Appl
C 22	36.2	7.4	126512	10	US-09-804-474A-3	Sequence 3, Appli
23	36	7.4	668	9	US-09-924-400-293	Sequence 293, App
24	36	7.4	668	9	US-10-012-896-367	Sequence 367, App
25	36	7.4	668	9	US-09-895-793-367	Sequence 367, App
26	36	7.4	668	9	US-09-895-814-367	Sequence 367, App
27	36	7.4	668	9	US-10-010-940-367	Sequence 367, App
28	36	7.4	668	10	US-09-759-143-367	Sequence 367, App
29	36	7.4	668	10	US-09-780-669-367	Sequence 367, App
30	36	7.4	668	10	US-09-810-936-293	Sequence 293, App
31	36	7.4	668	10	US-09-822-827-367	Sequence 367, App
32	36	7.4	668	10	US-09-429-755-293	Sequence 293, App
C 33	35.4	7.2	640681	10	US-09-790-988-1	Sequence 1, Appli
C 34	35	7.2	1569	10	US-09-849-566-9	Sequence 9, Appli
C 35	35	7.2	1569	10	US-09-907-859-9	Sequence 9, Appli
36	34.8	7.1	468	9	US-09-918-995-13121	Sequence 13121, A
37	34.8	7.1	515	9	US-09-796-692-7170	Sequence 7170, Ap
38	34.8	7.1	515	9	US-10-040-862-7170	Sequence 7170, Ap
39	34.8	7.1	5456	10	US-09-834-975-971	Sequence 971, App
C 40	34.8	7.1	1503841	9	US-09-946-807-1	Sequence 1, Appli
C 41	34.8	7.1	1503841	10	US-09-795-668-1	Sequence 1, Appli
C 42	34.8	7.1	1503841	10	US-09-795-668-1	Sequence 1, Appli
C 43	34.6	7.1	533	9	US-09-796-692-3419	Sequence 3419, Ap
C 44	34.6	7.1	533	9	US-10-040-862-3419	Sequence 3419, Ap
C 45	34.6	7.1	605	10	US-09-811-284-113	Sequence 113, App

ALIGNMENTS

RESULT 1

US-10-029-359A-3
; Sequence 3, Application US/10029359A
; Publication No. US20030100056A1
; GENERAL INFORMATION:
; APPLICANT: Attarsand, Anelli
; TITLE OF INVENTION: Protein Cluster II.
; FILE REFERENCE: 10806-156
; CURRENT APPLICATION NUMBER: US/10/029,359A
; CURRENT FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 489
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (55)..(489)
; OTHER INFORMATION:
US-10-029-359A-3

Query Match		100.0%	Score 489;	DB 9;	Length 489;
Best Local Similarity		100.0%	Pred. No. 8.2e-124;		
Matches 489;		Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	1	CATCTAAAGACATGATGAAGTTGATTCGCAAAATCTCTATCAATATACCAACACACA	60		
Db	1	CATCTAAAGACATGATGAAGTTGATTCGCAAAATCTCTATCAATATACCAACACACA	60		
QY	61	GGTGTCACTGACAACTCTATCAAAAGATGCAAAATCTCTTGTGGATTATGACAGACCT	120		
Db	61	GGTGTCACTGACAACTCTATCAAAAGATGCAAAATCTCTTGTGGATTATGACAGACCT	120		
QY	121	CAAAATCTAGAAAAAAGATCAGACGATCAGATACATATGCGCCCGGGTTGCAATAA	180		
Db	121	CAAAATCTAGAAAAAAGATCAGACGATCAGATACATATGCGCCCGGGTTGCAATAA	180		
QY	181	GCCTGTTTATGAGCTCAAAAGAAAAATTAAGACTTCTTCTACTGTGGATTACACACAA	240		
Db	181	GCCTGTTTATGAGCTCAAAAGAAAAATTAAGACTTCTTCTACTGTGGATTACACACAA	240		

Db 269 CCAANAAGACAGAAACCAACTGCTCTCGANAATAGTTANAGCNTTCTATATGCTCTG 210
QY 296 ACCATTATCTGGAATACATGGATTGGTTTCATGAAGCAGATGATATATATATATA 355
Db 209 ACCATTATCTGGAATACATGGANTGGTTTCATGAAGCAGATGATATATATATATA 150
QY 356 TCACATTTGGACAACTTGAATGGCTTCTCACAACATATAAACCTGATGAATCCACTTACT 415
Db 149 TCACATTTGGACAACTTGAATGGCTTCTCACAACATATAAACCTGATGAATCCACTTACT 90
QY 416 TTGGGAAAGATTTAAGCACTGCAGAAACAGAGCTACATGCTGGAGCAGGATATG 475
Db 89 TTGGGAAAGATTTAAGCNCCTGCAGAAACAGAGCTACATGCTGGAGNAGCAGGATATG 30
QY 476 TAC 478
Db 29 TAC 27

RESULT 4

US-10-235-056-12
; Sequence 12, Application US/10235056
; Publication No. US2003005915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; PRIOR FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1092
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-235-056-12

Query Match 62.8%; Score 307; DB 9; Length 1092;
Best Local Similarity 78.3%; Pred. No. 6.1e-74;
Matches 383; Conservative 0; Mismatches 100; Indels 6; Gaps 1;
QY 1 CATCTAAAGACGATGATGAAGTTGATTGCAAAATGCTAGTCAATATAATACCAACACA 60
Db 172 CATCTAGAAGACAAATGAACCTTCAATGCAGATTCTAGCCCAATAGATGAGAACACA 231
QY 61 GGTGCTACTGACAACTCTATCAAAAGATGAAAATCTTTGCTGGATTATGACAGACCT 120
Db 232 GACATTGCTGAAAACCTCTATCAGAAAAGTTAGAATCTTTGCTGGTTATGACCGCCCT 291
QY 121 CAAATCTAGAAAAAGATCAGACGATCAGAGATACATGCGGCCAGGGTTGCAATAAA 180
Db 292 CAAACCTAGAAAAAGCCCAACGCTCAAGCTACTTTGGGCCAGCGTTGTGAACAAA 351
QY 181 CGGTGTTTATGAGCTCAAAAGAAAATAAAGACTTCTCTACTGTGGGATTACACACAAA 240
Db 352 GTGTTGTTTATGAGTTTCAGAAAGAAAATAAAGACTTCCCTGCTGGGACTGAACACAAA 411
QY 241 GAAGACAGAACCAACTGCTCTGAAAATAGTTAAAGCTTTTCHATATGCTATGACCAT 300
Db 412 GAAGCAGAGATCAACTATATCTAGAAAACAAATTAAGCTTTTTCAGTATGTTATGAACAT 471
QY 301 TATCTGGAATACATGGATTGGTTTCATGAAGCAGATGATATATATATATATACACA 360
Db 472 TATTAGAAGTCTGATTGGTTTTCAGAAAGCAATTAAGCTTTTTCAGTATGTTATGAACAT 525
QY 361 TTGACAACTTGAATGGCTTCTCAACAACTATAAACCTGATGAATCCACTTACTTTGGG 420
Db 526 CTAGACAATTTGAGTGGCTTCTTCAAAATACGACCTTGAAGACCACTTACTTTGGG 585

QY 421 AAAAGATTAAAGCACTGTCAGAAAAACAGACTACATGCTGGAGGAGCAGGATATGACTG 480
Db 586 AGAAGATTAAAGCCCTTATGTAAAGCAGGCTACATGAGTGGAGGAGCAGGATATGACTA 645
QY 481 AGCAAAAGAA 489
Db 646 AGCAAAAGAA 654

RESULT 5

US-10-235-056-2
; Sequence 2, Application US/10235056
; Publication No. US2003005915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1794
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-235-056-2

Query Match 62.8%; Score 307; DB 9; Length 1794;
Best Local Similarity 78.3%; Pred. No. 7.7e-74;
Matches 383; Conservative 0; Mismatches 100; Indels 6; Gaps 1;
QY 1 CATCTAAAGACGATGATGAAGTTGATTGCAAAATGCTAGTCAATATAATACCAACACA 60
Db 234 CATCTAGAAGACAAATGAACCTTCAATGCAGATTCTAGCCCAATAGATGAGAACACA 293
QY 61 GGTGCTACTGACAACTCTATCAAAAGATGAAAATCTTTGCTGGATTATGACAGGACCT 120
Db 294 GACATTGCTGAAAACCTCTATCAGAAAAGTTAGAATCTTTGCTGGTTATGACCGCCCT 353
QY 121 CAAATCTAGAAAAAGATCAGACGATCAGAGATACATGCGGCCAGGGTTGCAATAAA 180
Db 354 CAAACCTAGAAAAAGCCCAACGCTCAAGCTACTTTGGGCCAGCGTTGTGAACAAA 413
QY 181 CGGTGTTTATGAGCTCAAAAGAAAATAAAGACTTCTCTACTGTGGGATTACACACAAA 240
Db 414 GTGTTGTTTATGAGTTTCAGAGAAAATAAAGACTTCCCTGCTGGGACTGAACACAAA 473
QY 241 GAAGACAGAACCAACTGCTCTGAAAATAGTTAAAGCTTTTCTATATGCTATGACCAT 300
Db 474 GAAGCAGAGATCAACTATATCTGAAAACAAATTAAGACTTTTTCAGTATGTTATGAACAT 533
QY 301 TATCTGGAATACATGGATTGGTTTCATCAAGCAGATGATATATATATATATACACA 360
Db 534 TATTAGAAGTCTGATTGGTTTTCAGAAAGCAATTAAGCTTTTTCAGTATGTTATGAACAT 587
QY 361 TTGACAACTTGAATGGCTTCTCAACAACTATAAACCTGATGAATCCACTTACTTTGGG 420
Db 588 CTAGACAATTTGAGTGGCTTCTTTCAAATAAGACCTTGAAGACCACTTACTTTGGG 647
QY 421 AAAAGATTAAAGCACTGCAGAAAAACAGGACTACATGCTGGAGGAGCAGGATATGACTG 480
Db 648 AGAAGATTAAAGCTTATGTAAAGCAGGCTTACATGAGTGGAGGAGCAGGATATGACTA 707
QY 481 AGCAAAAGAA 489
Db 708 AGCAAAAGAA 716

RESULT 6

US-10-029-359A-5
; Sequence 5, Application US/10029359A
; Publication No. US20030100056A1
; GENERAL INFORMATION:
; APPLICANT: Attersand, Anneli
; TITLE OF INVENTION: Protein Cluster II
; FILE REFERENCE: 10806-156
; CURRENT APPLICATION NUMBER: US/10/029,359A
; PRIORITY FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(934)
; OTHER INFORMATION:
US-10-029-359A-5

Query Match 62.5%; Score 305.4; DB 9; Length 1560;
Best Local Similarity 78.1%; Pred. No. 2e-73;
Matches 382; Conservative 0; Mismatches 101; Indels 6; Gaps 1;

QY	1	CATCTAAAGAGACTGATGAAGTTGATTCGAATGCTAGTCATCAATAATACAGAACACA	60
DB	17	CATCTAGAGGCAATGAACCTCAATGCAGATTCAGCAACATATAAGATGAGAACACA	76
QY	61	GGTGCTACTGACAACTCTATCAAAAGATGAAATTCCTTGGTGATTTATGACAGACCT	120
DB	77	GACATGCTGAAACCTCTATCAGAAAGTTAGAATTCCTTGGTGATTTATGACAGACCT	136
QY	121	CAAAATCTAGAAAAAGATCAGACGCATCAGAGATACATGGCCCGGTTGCAATAAA	180
DB	137	CAAAACCTAGAAAAAGGCCAACACGCTCAAGCTACTTGGCCCGGTTGTAACAAA	196
QY	181	GGTGTGTTTATGAGCTCAAAAGAAATAAGACTTCTCTACTGTGGGATTCACACACAA	240
DB	197	GTGTGTTTATGAGTTCAGAGAAATAAGACTTCTCTACTGTGGGATTCACACACAA	256
QY	241	GAAGACAGAAACCACTCTCTCGGAAATAGTAAAGCTTTCTATATGCTCATGACCAT	300
DB	257	GAAGCAGAGATCAACTACTCTGGAACAAATTAAGCTTTTTCAGTATGTTTCATGAACAT	316
QY	301	TATCTGGAATACATGGATGGTTCATGAAGCAGATGATATATATATATATATACACA	360
DB	317	TATTTACAAGATGCTGATTTGGTTTTTGAAGCAGATGATGA-----CACGTATGTCATA	370
QY	361	TTGGACAACCTGAATGGCTTCTCAAACTATACCCCTGATGAATCCACTTACTTTGGG	420
DB	371	CTAGACAATTTGAGTGGCTTCTTCAAAATACGACCTGAAGAACCCATTTACTTTGGG	430
QY	421	AAAGATTTAAGCACTGCAGAAAAAGCACTACATGCTGAGGAGCAGGATATGTTACTG	480
DB	431	AGAAGATTTAAGCCCTTATGTAAGCAGGCTACATGAGTGGAGGAGCAGGATATGTTACTA	490
QY	481	AGCAAGAA 489	
DB	491	AGCAAGAA 499	

RESULT 7

US-10-235-056-13
; Sequence 13, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630

; CURRENT APPLICATION NUMBER: US/10/235,056
; PRIORITY FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIORITY FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 1092
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-10-235-056-13

Query Match 61.5%; Score 300.6; DB 9; Length 1092;
Best Local Similarity 77.5%; Pred. No. 3.4e-72;
Matches 379; Conservative 0; Mismatches 104; Indels 6; Gaps 1;

QY	1	CATCTAAAGAGACTGATGAAGTTGATTCGAATGCTAGTCATCAATAATACAGAACACA	60
DB	172	CACCTCAAAGGACAGATGGACTTCAATGCAGATTCAGCAACATATAAGATGAGAACACA	231
QY	61	GGTGCTACTGACAACTCTATCAAAAGATGAAATTCCTTGGTGATTTATGACAGACCT	120
DB	232	GACGTGCTGAGAACCTCTATCAGAAAGTTAAAGTTCCTTGGTGATTTATGACAGACCT	291
QY	121	CAAAATCTAGAAAAAGATCAGACGCATCAGAGATACATGGCCCGGTTGCAATAAA	180
DB	292	CAAAATCTAGAAAAAGGCCAACACGCTCAAGCTACATGGCCCGGTTGTAATAAA	351
QY	181	GGTGTGTTTATGAGCTCAAAAGAAATAAGACTTCTCTACTGTGGGATTCACACACAA	240
DB	352	GTGTATTTATGAGTTCAGAAAGAAATAAGACTTCCCTACTGTGGGCTGGAACACAA	411
QY	241	GAAGACAGAAACCACTCTCTCGGAAATAGTAAAGCTTTCTATATGCTCATGACCAT	300
DB	412	GAAGCAGAGACCACTCTCTGGAACAAATTAAGCGTTTTCAGTATGCTCATGACCAT	471
QY	301	TATCTGGAATACATGGATGGTTCATGAAGCAGATGATATATATATATATATACACA	360
DB	472	TACTTAGAAGATGCTGACTGGTTTATGAAGCAGACGATGA-----CACCTATGTCATA	525
QY	361	TTGGACAACCTGAATGGCTTCTCAAACTATACCCCTGATGAATCCACTTACTTTGGG	420
DB	526	CTGGACAATCTGAGATGGCTTCTATCAAAATTAAGCTTAACCCCTGAACAGCCCATTTACTTTGGG	585
QY	421	AAAGATTTAAGCACTGCAGAAAAAGCACTACATGCTGAGGAGCAGGATATGTTACTG	480
DB	586	AGAAGATTTAAGCCCTTATGTAAGCAGGATACATGAGTGGAGGAGCAGGATATGTTCTA	645
QY	481	AGCAAGAA 489	
DB	646	AGCAAGAA 654	

RESULT 8

US-10-235-056-4
; Sequence 4, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; PRIORITY FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIORITY FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1440
; TYPE: DNA
; ORGANISM: Rattus norvegicus

US-10-235-056-4

Query Match	61.5%;	Score 300.6;	DB 9;	Length 1440;	
Best Local Similarity	77.5%;	Pred. No. 3.9e-72;			
Matches 379;	Conservative	0;	Mismatches 104;	Indels 6;	Gaps 1;
QY	1	CATCTAAAAGACTGATGAAGTTGATTGCAATGTAGTCATCATATAACAGAACACA	60		
Db					
325	CACCTCAAAGGACAGATGGACTTCAATGCAGATTCAGCCAACTATAGCCAACTAAGATGAGAACACA	384			
QY	61	GGTGCTACTGCAAAACTCTATCAAAGAGATGAAATTCCTTCTGGATTATGACAGGACCT	120		
Db					
385	GACGTGCTGGAACCTCTATCAGAAAGTTAAAGTTCCTTTGGGTTATGACAAGCCCT	444			
QY	121	CAAAATCTAGAAAAAAGATCAGACGCATCAGATACATGGGCCCGAGGTTGCAATAAA	180		
Db					
445	CAAAATCTAGAAAAAAGGCCAAACACGCTCAAAAGCTACATGGGCCCGAGGTTGTAATAAA	504			
QY	181	CGTTGTTTATGAGCTCAAAAGAAAAATAAGACTTCTCTACTGTGGGATACACACCAA	240		
Db					
505	GTGTTATTTATGAGTTCAGAAAGAAAAATAAGACTTCCTCTACTGTGGGCTGGAACCCAAA	564			
QY	241	GAAGACAGAAACCACCTGCTCTGGAAAATAGTTTAAAGCTTTCTATATGCTCATGACCAT	300		
Db					
565	GAAGCAGAGAACACTGTACTGGAACAACTTAAAGCGTTTCTCAGTACGTACATGACCAT	624			
QY	301	TATCTGGAAATACATGGATTGGTTTCATGAAAGCAGATGATATATCTATATATATCACA	360		
Db					
625	TACTTGAAGATGCTGACTGGTTTATGAAGCAGACGATGA-----CAGCTATGCTATA	678			
QY	361	TTGACACACTTGAATGGCTTCTCAGAACTATAACCTGATGAATCCACTTTACTTTGGG	420		
Db					
679	CTGACAATCTGAGATGGCTTCTATCAAAGTATAACCTCGAACAGCCCACTTTACTTTGGG	738			
QY	421	AAAAGATTTAAGCACTCCAGAAAAAGGACTACATGACTGGAGGACGAGATATGACTG	480		
Db					
739	AGAAGATTTAAGCCCTATGTGAAGCAGGGATACATGATGGAGGAGCAGGATATGCTTA	798			
QY	481	AGCAAAGAA	489		
Db					
799	AGCAAGGAA	807			

```

RESULT 9
US-10-235-056-14
; Sequence 14, Application US/102335056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE REFERENCE: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1092
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-235-056-14

```

```

Query Match      59.8%   Score 292.6; DB 9; Length 1092;
Rest Local Similarity 76.5%; Pred. No. 5.3e+70;
Matches 374; Conservative 0; Mismatches 109; Indels 6; Gaps 1;

QY    1 CATCTAAAGACGTGTAAGTGGATTGCCAATGCTAGTCATCATATAATACCAGAACACA 60
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     172 CACCTCAAAGCAGATGGAAGTTCAATGCAGATTCGACGCAACATATAAGATGAGAACAATA 231

```

QY	61	GGTGTCACTGACAAACTCTATCAAAAGATCAAAATCTTTGCTGGATTATGACAGGACCT	120
Db	232	CGCTTGGCTGAGAACCTCTATCAGAAGTTAAATTTCTTTGTTGGTTATGACAAGTCCT	291
QY	121	CAAAATCTAGAAAAAAGATCAGACCCATCAGAGATACATGGGCCCGCAGGTTGCAATAAA	180
Db	292	CAAAATCTAGAAAAAAGGCCAAACATGTCAAAGCTACGTGGGCCCGCAGGTTGTAATAAA	351
QY	181	GGCTTGTATTGAGCTCAAAAGAAAAATAAGACTTCTCTACTGTGGGATTACACACCAA	240
Db	352	GTGTTATTATCAGTTGCGAAGAAAATCAAGACTTCCCTACTGTGGGATTGAACACAA	411
QY	241	GAAACAGACAAACCACTGTCCTGGAATAAGTAAAGCTTTTCTATATGCTCATGACCAT	300
Db	412	GAAAGCAGAGCAACTATATTGGAACAATTAAGCTTCCAGTATCTACATGACCAT	471
QY	301	TATCTTGAATACATCGATTGGTTTCATGAAAGCAGATGATGATATATGTATATATATCACA	360
Db	472	TATTTAGAAGATGCTGACTGGTTTATGAAAGCAGATGACGA-----CACATACGTCATT	525
QY	361	TTGGACAACCTTGAATGGCTTCTCAAAACTATAACCCCTGATGAATCCACTTACTTTGGG	420
Db	526	GTGGACAACCTTGAGATGGCTTCTATCAAGTATAACCCCTGAACAACCCATTACTTTGGG	585
QY	421	AAAAGATTTAAGCACTGCGAAAAACAGGACTACATGCTGGAGGACGAGGATATGTAAGT	480
Db	586	CGAAGATTTAGCCCTATGTTGAAGCAGGATACATGAGCGGAGGACGGGTATGTCCTA	645
QY	481	AGCAAAAGAA	489
Db	646	AGCAAGGAA	654

```

RESULT 10
US-10-235-056-6
; Sequence 6, Application US/10235056
; Publication No. US20030059915A1
; GENERAL INFORMATION:
; APPLICANT: Canfield, William M.
; APPLICANT: Cummings, Richard D.
; APPLICANT: Ju, Tongzhong
; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
; FILE OF INVENTION: 5820.630
; CURRENT APPLICATION NUMBER: US/10/235,056
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/461,321
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1469
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-235-056-6

```

	Query Match	59.8%;	Score 292.6;	DB 9;	Length 1469;
	Best Local Similarity	76.5%;	Pred. No. 6e-70;		
	Matches 374;	Conservative 0;	Mismatches 109;	Indels 6;	Gaps 1
QY	1	CATCTAAAGACGTGATGACGTGATTGCAATGCTAGTCATCATATAATACCAGACACA	60		
Db	351	CACCTCAAGGACAGATGAATCTCAATGCCAGATCCAGGCCAACATAAAGATGAGAACATA	410		
QY	61	GGTGCTACTGCACAACTCTATCAAAAGATCAAAATTCCTTGTGCGATTATGACAGAGCT	120		
Db	411	GACGTGCTGAGAACCTCTATCAGAAGTTAAAATTCCTTGTGGGTTATGACAAAGTCT	470		
QY	121	CAAAATCTAGAAAAAAGATCAGACGCATCAGAGATACATGGGCCACGGTTGCAATAAA	180		
Db	471	CAAAATCTAGAAAAAAGGCCAACATGTCARAAGCTACGTGGGCCACGGTTGTAATAAA	530		
QY	181	CGCTGTGTTATGACCTCAAAGAAAAATAAAGACTTCTCTACTGTGGGATTACACACCAAA	240		

;; PRIOR FILING DATE: 1999-12-15
;; NUMBER OF SEQ ID NOS: 25
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 8
;; LENGTH: 1172
;; TYPE: DNA
;; ORGANISM: Caenorhabditis elegans
US-10-235-056-8

Query Match 21.58; Score 105; DB 9; Length 1172;

Best Local Similarity 55.28; Pred. No. 9.4e-19;
Matches 229; Conservative 0; Mismatches 180; Indels 6; Gaps 1;

QY 68 CTGACAAACTCTATCAAAAGATGAAATCTTTGCTGGATTATGACAGGACCTCAAAATC 127
DB 290 CCGAGGAAGTTGCGAAGAGGTTCCGCTCTCTGTTGGATTCTCACCGGAACAGATC 349
QY 128 TAGAAAAAGATGACAGCAGATGACAGATACATGAGGCGCCAGGTTGCAATAAAGCGTTGT 187
DB 350 ACGATAAAGCGGGAACACAGCTCAAGCCACCTGGGCCAAGCGGTGTAATAAGTACGTAT 409
QY 188 TTATGACCTCAAGAAATTAAGACTTCTCTACTGCTGGGATTACACACCAAAAGACACA 247
DB 410 TCATGTCATCAGAGAGGATGCGAAGCTCCAGCACTCCAGCCATCAACCTAAATGTCTCTGAAGGCA 469
QY 248 GAAACCAACTGCTGCTGGAATAATAGTTAAAGCTTTTCTATATGCTCATGACCATTTATCTGG 307
DB 470 GAGATTACTTGTGGCAAAACAAAGAGGACCATTCAAATACATTTATGACCATCACCTGA 529
QY 308 AATACATGGATTGGTTTCATGAAAGCAGATGATGATATATATATATATATATATATATGACAT 367
DB 530 ACGACTACGACTGGTTCTCTGAAAGCCGACGACGA-----TACCTATGTGGTGATGAA 583
QY 368 ACTTGAATGCTTCTCACAACTATAAACCCTGATGATCCACTTACTTTGGGAAGAT 427
DB 584 ATCTCGGTTTCATGCTATTGGCTATTCACCGGATGAGCCAAATTCACCTTTGGATGCAAGT 643
QY 428 TTAAGCACTGCAGAAACAGAGACTACATGATGAGGAGGAGGAGGATGATGATGAG 482
DB* 644 TTAAGCCATTACACAGGAGGATATCATATGTTGGAGTGGTGAGTATGTCCTCAG 698

RESULT 15

US-10-235-056-18
;; Sequence 18; Application US/10235056
;; Publication No. US20030059915A1
;; GENERAL INFORMATION:
;; APPLICANT: Canfield, William M.
;; APPLICANT: Cummings, Richard D.
;; APPLICANT: Ju, Tongzhong
;; TITLE OF INVENTION: CORE 1 3-GALACTOSYL TRANSFERASES
;; FILE REFERENCE: 5820.630
;; CURRENT APPLICATION NUMBER: US/10/235.056
;; CURRENT FILING DATE: 2002-09-04
;; PRIOR APPLICATION NUMBER: US 09/461,321
;; PRIOR FILING DATE: 1999-12-15
;; NUMBER OF SEQ ID NOS: 25
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 18
;; LENGTH: 1167
;; TYPE: DNA
;; ORGANISM: Drosophila melanogaster
US-10-235-056-18

Query Match 18.98; Score 92.2; DB 9; Length 1167;

Best Local Similarity 52.48; Pred. No. 3e-15;
Matches 230; Conservative 0; Mismatches 203; Indels 6; Gaps 1;

QY 50 ACCAGAACACAGGTGTCTACTGACAACTCTATCAAAAGATGAAATTTCTTTGCTGGATTA 109
DB 257 ACGAAGACTCGACCATTTGGCGGAGGACTGTACAGCGAGGTGCGTCTCTGCTGGATCA 316
QY 110 TGACAGGACCTCAAAATCTAGAAAAAAGATCAGACGCATCAGAGATACATGGGCCAGG 169

DB 317 TGACCAATCCGAGCAACCATCAGAAGAAGCGCGCCACGTCGAAAGCGCACCTGGGGCAAGC 376
QY 170 GTTGCATTAAGCGTTGTTTATGAGCTCAAAAGAAATTAAGACTTCTCTACTCTGGAT 229
DB 377 GTTGCATTAAGCGTTGTTTATGAGCTCAAAAGAAATTAAGACTTCTCTACTCTGGAT 436
QY 230 TACACACCAAAAGAGACAGAAACCAACTGTCCTGGAAATAGTTAAAGCTTTTCTATATG 289
DB 437 TGGCCGTAGCGAGGTCGCAACCACTATGGGCGAAGAGAGGCGCTTACAAATACA 496
QY 290 CTGATGACCATTAATCTGGAATACATGATGTTGTTTCATGAAAGCAGATGATATATGTA 349
DB 497 TCTACGAGCATCACATCAACGACGCGGACTGTTCTCTGAAAGCTGACGATGA-----CA 550
QY 350 TATATATACATTTGGACAACTTGAATGCTTCTCACAACATATAACCTGATGAATCCA 409
DB 551 CATACACGATAGTGGAGAACATCGGATACATGCTGTATCCGTACAGTCCGGGAAACTCCAG 610
QY 410 CTTACTTTGGGAAAAGATTAAAGCACTTGCAGAAAACAGGACTACATGACTGGAGGAGCAG 469
DB 611 TCTACTTCGCTCAAGTTCAAGCCGTACGTGAACAAGCTACATGTCCGGCGGTGCCG 670
QY 470 GATATGCTACTGAGCAAGA 488
DB 671 GCTACGTTCTCAGCCGGA 689

Search completed: June 28, 2003, 00:27:37

Job time : 129.594 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 23:34:12 ; Search time 1231.81 Seconds
(without alignments)
6429.241 Million cell updates/sec

Title: US-10-029-359A-3

Perfect score: 489

Sequence:

1 catctaaaagactgatgaa.....gatatgtactgagcaaaaga 489

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	311.4	63.7	481	12	BF443807
2	264.2	54.0	696	13	BI102329
3	260	53.2	637	10	BE336820
4	221.2	45.2	614	9	AJ449471
5	177.6	36.3	642	9	AJ447748
6	167.4	34.2	528	17	AQ230580
					HS-2033_A

7	163	33.3	505	14	N47633
8	144.6	29.6	1051	17	CNS018X7
c	143.4	29.3	594	12	BF471960
c	141.2	28.9	480	17	BI5559
11	131.8	27.0	606	12	BG077752
12	126.2	25.8	1088	17	CNS05RY8
13	124.2	25.4	915	14	BQ736824
14	111	22.7	582	13	BI508514
15	111	22.7	646	13	BI509237
16	108.6	22.2	618	17	AZ456919
17	98.4	20.1	316	12	BE993641
18	96.4	19.7	497	9	AL712318
19	92	18.8	501	9	AI546714
20	91.6	18.7	458	9	AI964359
21	89.4	18.3	651	13	BJ114209
22	87	17.8	677	10	AW773412
23	85.2	17.4	462	9	AI236007
24	81	16.6	471	9	AI059600
25	79.8	16.3	474	9	AI546718
26	73.4	15.0	257	10	BB602602
27	72	14.7	518	13	BI450723
28	70.6	14.4	557	14	BM714438
29	70.4	14.4	434	12	BF938171
30	70.2	14.4	328	12	BF158206
31	67.6	13.8	544	13	BI502334
32	66.4	13.6	563	13	BJ040104
33	64.8	13.3	506	13	BM533050
34	64.2	13.1	399	13	BM289607
c	64	13.1	689	9	AI816690
36	59	12.1	1101	17	CNS0082F
37	56.6	11.6	614	10	AW419812
38	55.4	11.3	352	14	R03773
39	55	11.2	141	14	T10488
40	55	11.2	570	9	AL698324
41	53.8	11.0	559	13	BI240060
42	52.4	10.7	941	17	CNS006VT
43	52.2	10.7	737	9	AI387726
44	52.2	10.7	797	17	CNS03JD9
c	51	10.4	1101	17	CNS014PC

ALIGNMENTS

RESULT 1
BF443807
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BF443807
261354 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
EST 01-DEC-2000

481 bp
Sus scrofa

PIG.

Sus scrofa

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 481)

Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keele, J.W.

Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine

Unpublished (2000)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCTCCAGTCACGACG
 Plate: 94 row: J column: 20
 Seq primer: ATTTAGTGACACTATAG.
 Location/Qualifiers
 1. .481
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 2Pig"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
 Library made from pooled tissue from testis, ovary,
 endometrium, hypothalamus, pituitary, and placenta."
 81 c 112 g 121 t
 167 a
 BASE COUNT
 ORIGIN

Query Match 63.7%; Score 311.4; DB 12; Length 481;
 Best Local Similarity 79.7%; Pred. No. 6.9e-67;
 Matches 382; Conservative 0; Mismatches 91; Indels 6; Gaps 1;

QY 11 GACTGATGAAGTTGATTGCTCAATAGCTAGTCATCAATAATACCAGACACAGAGTGTCTACTG 70
 Db 2 GACAGATAAACTTTGGTGCAGATGCTAGCCAAATAGATGAGACACAGACATCGCTG 61
 QY 71 ACAAATCTATCAAAAGAGTAAATTTCTTTGCTGGATTATGACAGACCTCAAAATCTAG 130
 Db 62 AAAACCTCTATCAGAAAGTTAAATTTCTTTGCTGGTAAATGACAGAGGCTCAAAATCTAG 121
 QY 131 AAAAAAGATCAGACGCATCAGAGATACATGCGCCAGGCTTGCATTAAGCGTCTTTTA 190
 Db 122 AGAAAAGGCCAAACATGTCAAAGCTACATGCGCCAGGCTTGTAAACAAGGTGTCTTTTA 181
 QY 191 TGAGCTCAAAAGAAAATAAGACTTCTCTACTGTGGGATTACACCAAGAAAGACAGAA 250
 Db 182 TGAGTTCAAGAAATAAAGACTTCCCTGCTGGGATTAAACACCAGAGAGCGAG 241
 QY 251 ACCAATCTGCTGGAATAATAGTTAAAGCTTTCTATATGCTCATGACCATTTATCTGGAAT 310
 Db 242 ACCAATCTGCTGGAATAATAGTTAAAGCTTTCTATATGCTCATGACCATTTATCTGGAAT 301
 QY 311 ACATGATTGGTTCATGAAGCAGATGATATATATATATATATATATATATATATATATAT 370
 Db 302 ATCCGATATGGTTTGAAGCAGATGATGA-----TAGGTATGTATCTAGATAACT 355
 QY 371 TGAATGGCTCTTCACAACTATAACCTCGATGAATCCACTTACTTTGGGAAAAGATTTA 430
 Db 356 TGAGATGGCTTCTTCAAAATAACACCCCTGAAGAACCCATTTACTTTGGAGAGATTTA 415
 QY 431 AGCACTGCAGAAAACAGGACTACATGCTGGAGGAGCAGGATATGCTAGCAGCAAGAA 489
 Db 416 AGCCCTATGTAAGCAGGGCTACATGAGCGGAGGTGCAGGGTATGCTAGTAAAGAA 474

RESULT 2
 BI102329
 LOCUS
 DEFINITION
 696 bp mRNA linear EST 26-JUN-2001
 5', mRNA sequence.
 BI102329
 BI102329.1 GI:14553222
 EST.
 SOURCE
 mouse mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 696)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaphs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM1112 row: j column: 15
 High quality sequence stop: 696.
 Location/Qualifiers
 1. .696
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:5040878"
 /clone_lib="NCI_CGAP_Kid14"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.75 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library. I"
 245 a 136 c 147 g 168 t
 BASE COUNT
 ORIGIN

Query Match 54.0%; Score 264.2; DB 13; Length 696;
 Best Local Similarity 75.5%; Pred. No. 3.6e-55;
 Matches 357; Conservative 0; Mismatches 108; Indels 8; Gaps 2;

QY 19 AACTGATTGCAAAATGCTAGTCATCAATAATACCAGAA--CACAGGTGTCACCTGACAAAC 76
 Db 1 AACTTCAATGCAGATTTCCAGCCCAACATAAAGATGAGACGATGACGTTGCTGCAACC 60
 QY 77 TCTATCAAAAGATGAAAATTTCTTTGCTGGATTATGACAGGACCTCAAAATCTAGAAAAA 136
 Db 61 TCTATCAGAAAGTTAAATTTCTTTGCTGGTTATGACAAGTCTCAAAATCTAGAGAAA 120
 QY 137 AGATCAGACCCATCAGAGATACATGCGCCAGGCTTGCAATAAGCGTTGTTTATGAGCT 196
 Db 121 AGCCCAAAATGTCAAAGCTACGTGGCCCGCGCTGTATAAAGTGTATTATGAGTT 180
 QY 197 CAAAGAAAATAAAGACTTCTCTACTGTGGGATTACACCAAGAAAGACAGAAACCAAC 256
 Db 181 CGAAGAAAATCAAGACTTCCCTACTGTGGGATTGAACCAAGAAAGGACAGAGCAAC 240
 QY 257 TGCTCTGGAATAATAGTTAAAGCTTTTCTATATGCTCATGACCATTTATCTGGAATACATG 316
 Db 241 TATATTGGAACAATAATAAAGCTTTCCAGTATGATGATGACCATTTATTTAGAAAGTCTG 300
 QY 317 ATTGGTTCATGAAGCAGATGATATATATATATATATATATATATATATATATATAT 376
 Db 301 ACTGGTTTATGAAGCAGATGACGA-----CATACGCTCATTTGGAAGAACTTGAGT 354
 QY 377 GGCTTCTCAAACTATAACCTCGATGAATCCACTTACTTTGGGAAAAGATTATTAAGCACT 436
 Db 355 GGCTTCTATCAAGTATAACCTCGAACAACCCATTTACTTCGGCGCAAGATTAAAGCCCT 414
 QY 437 GCAGAAAACAGGACTACATGCTGAGGAGCAGGATATGCTAGCAGCAAGAA 489
 Db 415 ATGTGAAGCAGGATACATGAGCGGAGGAGCGGCTATGCTCTAAGCAAGAA 467

RESULT 3
 BE336820
 LOCUS
 DEFINITION
 637 bp mRNA linear EST 14-JUL-2000
 ba97a03.y1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:2957836 5'
 similar to TR:095044 095044 WUGSC:H30733B09.1 PROTEIN ; contains
 Alu repetitive element;; mRNA sequence.
 BE336820
 BE336820.1 GI:9189205
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 637)


```

VERSION AJ447748.1 GI:20214969
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
JOURNAL Phasianinae; Gallus.
COMMENT 1 (bases 1 to 642)
Buerstedde,J.M.
Gallus gallus bursal lymphocyte EST
Unpublished (2002)
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.

FEATURES
source
1..642
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="1717r1"
/clone_lib="riken1"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/note="CB inbred strain"
BASE COUNT 174 a 135 c 172 g 158 t 3 others
ORIGIN
Query Match 36.3%; Score 177.6; DB 9; Length 642;
Best Local Similarity 71.7%; Pred. No. le-33;
Matches 231; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1 CATCTAAAGACTGTGAAGTTGATTGCAAAATCTAGTCATCATATAATACCAGACACA 60
Db 312 CAGCTGCAGACAAATGAATTTCAATCGGACTCTGGACGACATAGAGATGAGATAGA 371

QY 61 GGTGTCTACTGACAACTCTATCAAAAGATGAAAATCTTTTGGTGGATTATGACAGGACCT 120
Db 372 AACATTTCNGATGACTCTATGAGAAGGTGAAAATCTTTTGGTGGTTCATGACTGGGCCT 431

QY 121 CAAATCTAGAAAAGATCAGACGATCAGAGATACATGGGCCCGAGGTTTCAATAAA 180
Db 432 CAAATCTAGAGAAGAACGACACATGTTTAAAGCCACCTGGCCGCGTTTCAATAAA 491

QY 181 CGCTTGTGTTAGCTCAAAAGAAAATAAGACTTCTCTACTCTGGGATTACACACAAA 240
Db 492 ATACTTTTATGAGCTCTGAGGAATAAAGACTTNCACACGTTGGCTTAGAAACAAA 551

QY 241 GAAGACAGAAACCACTGCTCTCGAAAATAATGTTAAAGCTTTTCTATATGCTCATGACCAT 300
Db 552 GAAGCGAGGACCAACTTTACTGGAAGACTATAAAGCTTTTCAATATGTTATGACCAT 611

QY 301 TATCTGGAATACATGGATTGGT 322
Db 612 TATTTGATGATGCTGATTGGT 633

RESULT 6
A0230580 528 bp DNA linear GSS 26-SEP-1998
LOCUS HS_2033_AL_F01_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate-2033 Col-1 Row-K, DNA sequence.
ACCESSION A0230580
VERSION A0230580.1 GI:3655809
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 528)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel.: (206) 616-3618
Fax: (206) 616-3887
Email: jwallaceu.washington.edu
Sequence Tagged Connector
Plate: 2033 row: K column: 1
Class: BAC ends
High quality sequence stop: 528.
Location/Qualifiers
1..528
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=2033 Col=1 Row=K"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 163 a 104 c 112 g 147 t 2 others
ORIGIN
Query Match 34.2%; Score 167.4; DB 17; Length 528;
Best Local Similarity 85.7%; Pred. No. 3.3e-31;
Matches 186; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 273 TAAAGCTTTTCTATATGCTCATGACCATTTATCTGGAATACATGATTTGTTTCATGAAAGC 332
Db 46 TCAGGCTCACTATGCTCATGACCATTTATCTGTAATACCTGGTTCGTGAGCTC 105

QY 333 AGATGATGATATGATATATATATATACATTTGGAACACTTGAATGGCTTCTTCAACAATA 392
Db 106 TGATGATGAACCATGTATATATATACCATTTGGAAACTCGAAATGGCTTATCACAATA 165

QY 393 TAACCTCTGATGATTCACCTTACTTTGGAAACATTTAAGCACTGCAGAAACAGGACTA 452
Db 166 TAACCATGATGAATACATCTACTTTGGAAAGATCTAACCGCTGCAGACTACAGGACTA 225

QY 453 CATGACTGGAGGAGGAGGATATGATCTAGCAAGAA 489
Db 226 CATGACTGGAGGAGGATATGATCTGCTGAGCACTGAA 262

RESULT 7
N47633 505 bp mRNA linear EST 14-FEB-1996
LOCUS yy54c02.rl Soares_multiple_sclerosis_2NbHMSp Homo sapiens cDNA
DEFINITION clone IMAGE:277346 5', mRNA sequence.
ACCESSION N47633
VERSION N47633.1 GI:1188799
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 505)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohldmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

```



```

VERSION BF471960.1 GI:11541143
SOURCE EST.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Bonaio, M.F., Lennon, G. and Soares, M.B.
JOURNAL Normalization and subtraction: two approaches to facilitate gene
MEDLINE Genome Res. 6 (9), 791-806 (1996)
COMMENT 97044477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEst@mail.nih.gov
cDNA Library Preparation: M.B. Soares Lab Clone Distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements. The following repetitive elements were
found in this cDNA sequence: 565-594, >GC-rich#Low_complexity
Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source           1..594
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH3-awu-c-04-0-UI"
                     /clone_lib="NIH_BMAP_M_S4"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_M_S4 library is a subtracted library of a series,
ultimately derived from a mixture of individually tagged
normalized libraries from ten regions of the mouse brain
(cerebellum, brain stems, olfactory bulbs, hypothalamus,
cortex, amygdala, basal ganglia, pineal gland, striatum,
hippocampus) after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: NIH_BMAP_M_S4,
NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
(NIH_BMAP_M_S4) was constructed as follows: PCR amplified
cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
was used as a driver in a hybridization with a pool of
the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
libraries in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library)
was purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
NIH_BMAP_M_S4 library. This procedure has been previously
described (Bonaio, Lennon and Soares, Genome Research
6:791-806, 1996)".
BASE COUNT      158 a 136 c 148 g 152 t
ORIGIN
Query Match      29.3%; Score 143.4; DB 12; Length 594;
Best Local Similarity 77.3%; Pred. No. 2.9e-25;
Matches 174; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 1 CATCTAAAGAGCTGATGAGTTCATGCAAACTGCTAGTTCATCATTAATACCAACACA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 235 CACCTCAAGAGGAGATGATGAACTTCATGCGAGATCCAGCAACATAAAGATGAGACA 176
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

61 GGTGTCACTGACAAACTCTATCAAAAGATGAAATCTTCTGCTGGATTATGACAGACCT 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175 GACGTTCTGAGACCTCTATCAGAAAGTTAAATCTTTTGGTTATGACAAGTCCT 116
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 CAAATCTAGAAAAGATCAGACGATCAGAGATACATGGGCCAGGTTTGCATATAA 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 CAAATCTAGAAAAGAGGCCAACATGTCAAAGTACGTGGGCCAGCGTTGTAATAAA 56
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 GCGTTGTTTATGAGCTCAAAAGAAAATAAAGACTTCTCTACTGTG 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 GTGTTATTATGAGTTCGGAAGAAAATCAAGACTTCCTCTCTTTG 11
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
LOCUS B15559
DEFINITION 345H11.TP CIT978SK1 Homo sapiens genomic clone A-345H11, DNA
sequence.
ACCESSION B15559
VERSION B15559.1 GI:2123308
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Adams, M.D., Kelley, J.M., Rounsley, S.R. and Venter, J.C.
JOURNAL Use of a BAC End Sequence Database for Sequence-Ready Map Building
COMMENT Unpublished (1997)
Other_GSSs: 345H11.TV
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..480
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="A-345H11"
                     /clone_lib="CIT978SK1"
                     /sex="Female"
                     /cell_type="Fibroblast"
                     /note="Vector: pBAC108L; Site 1: HindIII; Site 2: HindIII;
Caltech Human BAC Library A1"
BASE COUNT      144 a 90 c 75 g 171 t
ORIGIN
Query Match      28.9%; Score 141.2; DB 17; Length 480;
Best Local Similarity 78.0%; Pred. No. 9.9e-25;
Matches 170; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
QY 46 AAATACCAGAACACAGGTGTCACTGACAAAACCTCTATCAAAAGATGAAATTCCTTGTG 105
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 218 ATAGATGAGAACACAGACATTTGCTGAAACCTCTATCAGAAAGTTAGAAATCTTTGCTGG 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 106 ATTATGACAGACCTCAAAATCTAGAAAAAAGATCAGACGCATCAGAGATACATGGGCC 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 158 GTTATGACCGCCCTCAAAACCTTAGAGAAAAAGCCCAACACCGTCAAGTACTTGGGCC 99
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 CAGGTTGCAATAAAGGTTGTTTATGAGCTCAAAAGAAAATAAAGACTTCTCTACTGTG 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 98 CAGGTTGTAAACAAGTGTGTTTATGAGTTTCAGAGAGAAAATAAAGACTTCCCTGCTGTG 39
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 226 GGATTACACACCAGAACAGACAAAACCAACTGTCTCTG 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY	QY	DB	QY	DB
233	ACACCAAGACAGACAGAACCACTGTCCTCGAATAAGTTTAAAGTTTTCTATATGCT	190	CGGACGAGCTGTACAANNNGTCGCTGTTCTTTGCTGGTGATGACGGGACCCCAACCC	249
291		191		250
292		192		251
293		193		252
294		194		253
295		195		254
296		196		255
297		197		256
298		198		257
299		199		258
300		200		259
301		201		260
302		202		261
303		203		262
304		204		263
305		205		264
306		206		265
307		207		266
308		208		267
309		209		268
310		210		269
311		211		270
312		212		271
313		213		272
314		214		273
315		215		274
316		216		275
317		217		276
318		218		277
319		219		278
320		220		279
321		221		280
322		222		281
323		223		282
324		224		283
325		225		284
326		226		285
327		227		286
328		228		287
329		229		288
330		230		289
331		231		290
332		232		291
333		233		292
334		234		293
335		235		294
336		236		295
337		237		296
338		238		297
339		239		298
340		240		299
341		241		300
342		242		301
343		243		302
344		244		303
345		245		304
346		246		305
347		247		306
348		248		307
349		249		308
350		250		309
351		251		310
352		252		311
353		253		312
354		254		313
355		255		314
356		256		315
357		257		316
358		258		317
359		259		318
360		260		319
361		261		320
362		262		321
363		263		322
364		264		323
365		265		324
366		266		325
367		267		326
368		268		327
369		269		

.128	TAGAAAAAAGATCAGACGCATCAGACATACATGGGCCACGGTTGCAATAAACGGTTGT	187
250	TGAGACCAACAGCGCGCACGTCAAANNACCCTGGACGCGCCACTGCAACAATTGTGGTTT	309
188	TTATGAGCTCAAAGAATAAAGACATTCTACTGTGGGATTACACACCAAAAAGACA	247
310	TCATGAGCTCGTGAGAGCCCCAAC TTC CCACCGTGGGCGCTGGGTACGAAGGAGGGGC	369
248	GAAACCAACTGTCTCTGGAAAAATAGTTAAAGCTTTTCTATATGCTCATGACCAATTATCTGG	307
370	GGGACCAGCTGTCTCTGGAANCCATCCGGGSCCTCCACTAGCCCTACGAGCACCACGFCG	429
308	AATACATGGATTGGTTTCATGAAGCAGATGATATATGTATATATATACATTTGGACA	367
430	ACGAGGCCGATTTGGTTCTCTCAAGCGGACGATGA-----CACCTATGTAGTAGTGGACA	483
368	ACTTTGAAATGGCTTCTCACAAACTATAACCTGTATGAATCCACTTACTTTTCGGAAAAGAT	427
484	ACCTTCGCTGGTTCCTCGCCAAACCACACGCCCAGCGCCCATCTNCTTCGCGCCGAG-T	542
428	TTAAGCACTCGCAAAAACAGGACTTACATGACTGGAGGAGCAGGATATGTACTTGAGCAAG	487
543	TCAAACCTNCACCAAGCAGGGCTACATGTGCGCGCGCGGGCTACGCTTTGAGCTAAG	602
488 A	488	
603 A	603	
Db		

RESULT 13	BQ736824	LOCUS	AGENCOURT_8115982 NICHDC XGC Emb4	915 bp	mrna	linear	EST 16-JUL-2002
DEFINITION			IMAGE:5571880 5', mRNA sequence.				Xenopus laevis cDNA clone
ACCESSION	BQ736824		BQ736824.1	GI:21875721			
KEYWORDS			EST.				
SOURCE			African clawed frog.				
ORGANISM			Xenopus laevis				
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
			Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;				
			Xenopodinae; Xenopus.				
			1 (bases 1 to 915)				
REFERENCE			NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap .				
AUTHORS			National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
TITLE			Tumor Gene Index				
JOURNAL			Unpublished (1997)				
COMMENT			Contact: Robert Strausberg, Ph.D.				
			Email: cgapbs-r@mail.nih.gov				
			Tissue Procurement: Dr. Igor Dawid				
			cDNA Library Preparation: Life Technologies, Inc.				
			cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
			DNA Sequencing by: Agencourt Bioscience Corporation				
			Clone distribution: NCI-CGAP clone distribution information can be				
			found through the I.M.A.G.E. Consortium/LLNL at:				
			www.bio.llnl.gov/bbrp/image/image.html				
			Plate: LLAM12316	row: g	column: 17		
			High quality sequence stop: 656.				

BASE COUNT	262 a	217 c	188 g	248 t
ORIGIN				

Query Match	25.4%	Score 124.2	DB 14	Length 915
Best Local Similarity	60.3%	Pred. No. 1.8e-20		
Matches 243	Conservative 0	Mismatches 153	Indels 7	Gaps 2
QY	24	GATTGCAAAATGCTAGTCATCATATAAATACCCAGACACAGAGTGTCTACCTGACACAACTCTATCA	83	
Db	428	GGTTGGAATGTCCTCATTTTCATCTGCCAGAGATATATAGTGTGAGTGAGGAACCTCTCCAA	487	
QY	84	AAAGATGAAATTTCTTTGCTGGATTATGACAGGACCTCAAAATCTAGAAAAAAGATCAG	143	
Db	488	AAAGGTTCAGAGTGCCTTTGTTGGATCATGACTGACCCGACACAACTTAAGACCAACGAT	547	
QY	144	ACGCATCAGAGATACATGGCCCGAGGTTGCATAAAGCGTTGTTTATGAGCTCAAAAGA	203	
Db	548	TCATGTGAAGATTTCTCTGGACCCGTCACCTGCAATGTAGCGCTATTATGAGCTCCATCAC	607	
QY	204	AAATAAAGACTTCTCTACTGTGGGATTACACACCAAGAGACAGACAACTGTCCTG	263	
Db	608	CGATGAAGACTTCCCGCCCATTTGCTCTGGAACTGGGGAAGCGACAGACAACTTTACTG	667	
QY	264	GAAATAGTTAAAGCTTTTCTATATGCTCATGACCACTTATCTGGAATACA-TGGATTGGT	322	
Db	668	GAAACAATTCGGTCCCTTCCATTTACGCCATAAGTACTACCTGAATGAGACAGAATGGGT	727	
QY	323	TCATGAAGACGATGATATATATGATATATATATATATATGATGACAACTTGGCAACTTGC	382	
Db	728	CTTCTCAAGACGATGATGA-----TACATATGTAATGATGGAATTTACGATGGATGC	781	
QY	383	TCACAACTATACCTCTGATGAATCCACTTACTTTGGGAAAAG	425	
Db	782	TTTCGAACATATACCGCTGACCACCTATTTTACTTTGGGAAAAG	824	
RESULT 14				
BI508514				
LOCUS		582 bp	mrna	linear
DEFINITION		BB170009B21H02.5	Bee Brain Normalized/Subtracted Library, BB17 Apis	EST 08-APR-2002
ACCESSION		BI508514		
VERSION		BI508514.1	GI:15358888	
KEYWORDS				
SOURCE			honeybee.	
ORGANISM			Apis mellifera	
REFERENCE			Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.	
AUTHORS			1 (bases 1 to 582) Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L., Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.	
TITLE			Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee	
JOURNAL			Genome Res. 12 (4), 555-566 (2002)	
MEDLINE			21929762	
COMMENT			Contact: Gene E. Robinson Department of Entomology University of Illinois 505 S. Goodwin Ave., Urbana, IL 61801, USA Tel: 217 265 0309 Fax: 217 244 3499 Email: generobi@life.uiuc.edu This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield. PCR Primers FORWARD: TAATACGACTCACTATAGG BACKWARD: ATTAACCTCACTAAAG Plate: BB170009B21 ROW: H column: 02 Seq primer: ACGGATACAAATTCACACAGA High quality sequence stop: 582. Location/Qualifiers 1. .582	
FEATURES				
source				

```

/organism="Apis mellifera"
/strain="mixed strains of European bees, predominantly
A.m. ligustica"
/db_xref="taxon:7460"
/clone="BB170009B21H02"
/clone.lib="Bee Brain Normalized/Subtracted Library, BB17"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/organism="Apis mellifera"
/strain="mixed strains of European bees, predominantly
A.m. ligustica"
/db_xref="taxon:7460"
/clone.lib="Bee Brain Normalized/Subtracted Library, BB17"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/organism="Apis mellifera"
/strain="mixed strains of European bees, predominantly
A.m. ligustica"
/db_xref="taxon:7460"
/clone.lib="Bee Brain Normalized/Subtracted Library, BB17"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"

```

```

Query Match 22.7%; Score 111; DB 13; Length 582;
Best Local Similarity 60.7%; Pred. No. 3.1e-17;
Matches 202; Conservative 0; Mismatches 125; Indels 6; Gaps 1;

QY 157 ACATGGCCCCAGGTTGCAATAAAGCGTTGTTATGAGCTCAAAAGAAAATAAAGACTTC 216
D 157 ACATGGCCCCAGGTTGCAATAAAGCGTTGTTATGAGCTCAAAAGAAAATAAAGACTTC 216
D 2 ACATGGGAAAACGTTGCAATATCTTTTATGAGTCTGCTGAGATACAAATTA 61
QY 217 TCTACTGTGGATTTACACCAACCAAGACAGAACCAACTGCTCTGGAATAATAGTTAA 276
D 217 TCTACTGTGGATTTACACCAACCAAGACAGAACCAACTGCTCTGGAATAATAGTTAA 276
D 62 CCTACAGTTGTTTATCAGTAAAGAGAGAGATAATTTATGGGCAAAACCAAGAA 121
QY 277 GCTTTCTATATGCTATATATATATATATATATATATATATATATATATATATATATAT 396
D 277 GCTTTCTATATGCTATATATATATATATATATATATATATATATATATATATATATAT 396
D 182 GATGA-----TACATATATGCTTATGAGTAAAGTACAAAGTATGTTGTTATGAAAGCTGAT 235
QY 397 CCTGATGATCCACTTACATATATATATATATATATATATATATATATATATATATATAT 456
D 397 CCTGATGATCCACTTACATATATATATATATATATATATATATATATATATATATATAT 456
D 236 TCAAAATCGCCATTTATATATATATATATATATATATATATATATATATATATATATAT 295
QY 457 ACTGAGGAGGAGGATATGCTAGTAAAGAA 489
D 457 ACTGAGGAGGAGGATATGCTAGTAAAGAA 489
D 296 ACTGAGGAGGAGGATATGCTAGTAAAGAA 328

```

```

RESULT 15
BI509237
LOCUS BB170009B20H02.5 Bee Brain Normalized/Subtracted Library, BB17 Apis
DEFINITION mellifera cDNA clone BB170009B20H02 5', mRNA sequence.
ACCESSION BI509237
VERSION BI509237.1 GI:15359611
KEYWORDS EST.
SOURCE honeybee.
ORGANISM Apis mellifera
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
Apoidea; Apis.
REFERENCE 1 (bases 1 to 646)
AUTHORS Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
TITLE Annotated expressed sequence tags and cDNA microarrays for studies
of brain and behavior in the honey bee
JOURNAL Genome Res. 12 (4), 555-566 (2002)
MEDLINE 2192762
COMMENT Contact: Gene E. Robinson
Department of Entomology

```

```

University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical
Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
Award in Functional Genomics to G.E. Robinson and an NSF
Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR Primers
FORWARD: TAATGAGCTCACTATAGGG
BACKWARD: ATTACCCCTCACTAAAG
Plate: BB170009B20 row: H column: 02
Seq primer: ACCGATAACAAATTCACACAGGA
High quality sequence stop: 646.
Location/Qualifiers
1. 646
/organism="Apis mellifera"
/strain="mixed strains of European bees, predominantly
A.m. ligustica"
/db_xref="taxon:7460"
/clone.lib="Bee Brain Normalized/Subtracted Library, BB17"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/organism="Apis mellifera"
/strain="mixed strains of European bees, predominantly
A.m. ligustica"
/db_xref="taxon:7460"
/clone.lib="Bee Brain Normalized/Subtracted Library, BB17"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"

```

```

BASE COUNT 217 a 84 c 139 g 204 t
ORIGIN
Query Match 22.7%; Score 111; DB 13; Length 646;
Best Local Similarity 60.7%; Pred. No. 3.1e-17;
Matches 202; Conservative 0; Mismatches 125; Indels 6; Gaps 1;

```

```

QY 157 ACATGGCCCCAGGTTGCAATAAAGCGTTGTTATGAGCTCAAAAGAAAATAAAGACTTC 216
D 157 ACATGGCCCCAGGTTGCAATAAAGCGTTGTTATGAGCTCAAAAGAAAATAAAGACTTC 216
D 2 ACATGGGAAAACGTTGCAATATCTTTTATGAGTCTGCTGAGATACAAATTA 61
QY 217 TCTACTGTGGATTTACACCAACCAAGACAGAACCAACTGCTCTGGAATAATAGTTAA 276
D 217 TCTACTGTGGATTTACACCAACCAAGACAGAACCAACTGCTCTGGAATAATAGTTAA 276
D 62 CCTACAGTTGTTTATCAGTAAAGAGAGAGATAATTTATGGGCAAAACCAAGAA 121
QY 277 GCTTTCTATATGCTATATATATATATATATATATATATATATATATATATATATATAT 396
D 277 GCTTTCTATATGCTATATATATATATATATATATATATATATATATATATATATATAT 396
D 182 GATGA-----TACATATATGCTTATGAGTAAAGTACAAAGTATGTTGTTATGAAAGCTGAT 235
QY 397 CCTGATGATCCACTTACATATATATATATATATATATATATATATATATATATATATAT 456
D 397 CCTGATGATCCACTTACATATATATATATATATATATATATATATATATATATATATAT 456
D 236 TCAAAATCGCCATTTATATATATATATATATATATATATATATATATATATATATATAT 295
QY 457 ACTGAGGAGGAGGATATGCTAGTAAAGAA 489
D 457 ACTGAGGAGGAGGATATGCTAGTAAAGAA 489
D 296 ACTGAGGAGGAGGATATGCTAGTAAAGAA 328

```

Search completed: June 28, 2003, 01:00:57
Job time : 1235.81 secs

